DOCUMENT RESUME

ED 447 623 EC 308 112

AUTHOR	Kornhaber, Mindy Laura
TITLE	Seeking Strengths: Equitable Identification for Gifted Education and the Theory of Multiple Intelligences.
PUB DATE	1997-00-00
NOTE	340p.; Ed.D. Thesis, Harvard University.
PUB TYPE	Dissertations/Theses - Doctoral Dissertations (041)
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	*Ability Identification; Economically Disadvantaged; Elementary Secondary Education; *Equal Education; *Gifted Disadvantaged; Minority Groups; *Multiple Intelligences; Program Effectiveness; Program Evaluation; *Talent Identification
IDENTIFIERS	Charlotte Mecklenburg Public Schools NC; Gardner (Howard); Montgomery County Public Schools MD; University of Arizona

ABSTRACT

This study examined three federally supported programs that utilize Howard Gardner's theory of multiple intelligences in the identification of giftedness in economically disadvantaged and minority group youth. Following an extensive review of the literature, three chapters examine each program in detail. Each chapter first sets the identification effort in its theoretical, historical, and community contexts, then describes and analyzes the assessment in light of eight conditions. The first program is the DISCOVER Project, a collaboration of the University of Arizona with nine local schools and districts that utilizes a continuum of problem types. The second program is the Problem Solving Assessment used in the Charlotte-Mecklenburg (North Carolina) schools for identifying children for gifted services. The third program is the Early Childhood Gifted Model Program in Montgomery County, Maryland, which emphasizes teaches' observations and efforts to elicit and develop students' intelligence. Evaluation of all three programs concluded that although all of the programs were identifying more underserved students as gifted, it was difficult to associate the improved rates of identification with the assessment procedures since none are actually drawing heavily from the theory of multiple intelligences. Eleven appendices provide supporting forms and other information for each of the three programs. (Contains approximately 200 references.) (DB)

Seeking Strengths:
Equitable Identification for Gifted Education and the Theory of Multiple Intelligences

Mindy Laura Kornhaber

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

1997

For my daughter, Paulina.

Only so much can be captured in words.

ACKNOWLEDGEMENTS

During my studies at Harvard I have been fortunate to work with many extraordinary teachers, including all three of my readers. I am especially grateful to Howard Gardner for encouraging me in so many ways and for his ongoing mentorship. Barbara Neufeld has helped me to understand how to balance the demands of scholarship and life. Gary Orfield has been the kindest, most accessible, and yet most rigorous adviser imaginable. I am deeply indebted to all of them for sharing their wisdom with me.

In the three sites that I visited for this investigation, I received invaluable and generous help. I am grateful to Carol Hylton, Carol Reid, Judy Rogers, and Brenda Romanoff for their intellectual and moral support throughout this process. Brian Bartels forwarded much-needed documents. June Maker, Waveline Starnes, and Anne Udall were kind enough to grant me access to their work and provided thoughtful responses to questions and drafts.

A number of fellow students have provided advice and guidance. I am indebted to Edward Miech, a doctoral student who freely spent several hours introducing me to the software I needed to code and sort my data. I very much appreciate the kindness of fellow doctoral students, friends, and colleagues at Harvard Project Zero. It is inspiring to be in the company of so many thoughtful and hard-working people.

The funding of this research was partly supported by awards from the Harvard Graduate School of Education and Radcliffe Grants for Graduate Women. It would not have been possible for me to graduate this year without this financial assistance.

My family is due very special recognition. I am exceedingly grateful to my parents for their love and for having shaped my approaches to thinking and questioning. My sister-in-law, Diane Wake, cheerfully pitched in during my site visits and other hectic times. My husband, Warren Wake, has tried to help me emulate his own speedy progress through the dissertation process by taking over many responsibilities on the home front. I appreciate all of his efforts. Paulina, our daughter, has endured a mother who has worked too much and played too little. I hope it is not just long hours of hard work, but a desire to learn and to understand, that she remembers from this experience.

TABLE OF CONTENTS

Chapter 1
IDENTIFYING UNDERREPRESENTED GIFTED YOUNGSTERS: ISSUES AND METHODS

INTRODUCTION

If it is a reasonable goal to meet the educational needs of all children, then it is reasonable to provide services to nurture gifted and talented children. However, these youngsters are chronically underserved by public schools (U.S. Department of Education, 1993). This is especially true for low-income and minority students, especially African American, American Indian, and Hispanic youngsters (Adams & Callahan, 1994; Borland, 1989; Callahan & McIntire, 1994; Ford, 1994, 1995; Frasier, 1989a, 1989b; Frasier, Garcia, & Passow, 1995; Harris & Ford, 1991; Hartley, 1991; Maker & Schiever, 1989; McDaniel, 1988; Schmidt, 1993; Swisher & Tonemah, 1991; Tannenbaum, 1983; Tonemah, 1991; U.S. Department of Education, 1993). Commonly, the proportion of African American and Hispanic students in gifted education is not even half of that in the wider school population (Kitano & Kirby, 1985; Perrine, 1989; U.S. Department of Education, 1993). The proportion of American Indians appears to be only one-fourth or less (Callahan & McIntire, 1994). The problem is so widespread that gifted programs have been described as "the most segregated educational programs in the United States" (Ford, 1995, p. 52).

Current identification practices are widely regarded as a major barrier to participation by poor and minority youngsters in gifted education. Critics assert that prevalent identification procedures for gifted education fail to detect the existing or potential strengths of those whose language, culture, or relationship to schooling differ from that of middle class white students (Adams & Callahan, 1994; Ford, 1994, 1995; Frasier, 1989a, 1989b, Frasier, Garcia, & Passow, 1995; Harris & Ford, 1991; Maker, 1992; Pfeiffer, 1989; Schmidt, 1993; Swisher & Tonemah, 1991; Torrance, 1978; U.S. Department of Education, 1993).

Alongside criticisms for inequitable identification, programs for high-ability youngsters are reproached on educational grounds. Opponents claim that grouping of high ability students does not provide marked benefits to bright youngsters and detracts from the learning of others (e.g., Oakes, 1985; Slavin, 1996). Much of the controversy about tracking, according to Kulik (1992), stems from comparisons of academic gains experienced by heterogeneously grouped students with ability-grouped students who are following the regular or common curriculum (rather than a differentiated curriculum). Using a common curriculum, the gains of high ability youngsters are not appreciable, whereas the gains for lower or middle tracked students are about the same as they would be in mixed-ability classrooms (Kulik, 1992). Relatedly, there is tension between those who claim all youngsters benefit in cooperative learning groups (e.g., Slavin, 1996), and their opponents, who claim that able youngsters spend much of their time tutoring their classmates at the expense of their own learning (Gallagher, 1994; Renzulli & Reis, 1991; Rogers, 1991). Again, different bases of comparison -- the kind of cooperative learning, the learning in mixed-ability groups versus the learning in other grouping arrangements -- lead to different claims.

One clear finding from various meta-analyses is that highly able youngsters learn more in programs that offer enriched or accelerated curriculum (Kulik, 1992; Kulik & Kulik, 1991; Rogers, 1991). These programs may take a variety of forms, including separate classrooms for identified youngsters, clustering of identified youngsters within regular classrooms, pull-out programs, or cross-grade grouping in particular subject areas.

A meta-analysis by Kulik and Kulik (1991), which included comparisons of youngsters taught in homogeneous and heterogenous classrooms, reveals that high-aptitude youngsters experience positive academic benefits of moderate size "in programs that are specially designed for gifted students" (Kulik & Kulik, p. 191). Based on a best-evidence analysis of 13 syntheses of ability grouping (including the Kuliks'), Rogers (1991) claims that when gifted youngsters are provided enriched or accelerated curriculum, they experience substantial academic gains. Her analysis leads her to assert:

> it is very clear that the academic effects of a variety of long and short-term grouping options for both the purposes of enrichment and acceleration are extremely beneficial for students who are academically or intellectually gifted or talented. There is no body of evidence that "the research says" otherwise! (Rogers, 1991, pp. 25-26).

Along with academic gains, youngsters in such enrichment programs may also experience social benefits, though the impact of such benefits is less well documented. Peter Rosenstein (personal communication, 1997), the executive director of the National Association for Gifted Children, argues that such programs keep bright youngsters engaged in school and prevent them from becoming drop-outs and disciplinary problems. Research on dropping out offers some support of this point. For example, Fine (1991) believed that many inner city dropouts she studied possessed greater ability than the youngsters who remained in school.

Among the obvious problems with such grouping practices is that they benefit not only bright youngsters (and thereby collide with our culture's anti-intellectual and egalitarian tendencies), but that they typically benefit primarily bright youngsters from already advantaged groups. The underrepresentation of poor and minority youngsters in such programs exacerbates existing educational inequities.

In 1988, to enhance gifted education in general and to provide greater access to youngsters traditionally underrepresented in such programs, Congress established the Jacob K. Javits Gifted Talented Students Education Program under Title IV, Part B of the Hawkins-Stafford Elementary and Secondary Amendments of 1988. The Javits Program was reauthorized in 1994 as part of the Improving America's Schools Act, under Title X, Part B. The legislation calls upon the Javits Program to foster a national focus on the needs of gifted and talented youngsters and to build national capacity for meeting those needs. The Javits Program does so partly through funding the National Research Center on Gifted and Talented, which supports and disseminates research related to gifted and talented students. In addition, the Javits Program provides grants to "state and local education agencies, institutions of higher education, and other public and private agencies ... to meet the needs of talented and gifted students" (U.S. Department of Education, 1994, p. 1). By law, half the grants awarded under the Javits Program must serve the needs of economically disadvantaged students. The grants are also supposed to favor programs with a state-wide or regional emphasis (U.S. Department of Education, 1994; U.S. Congress, 1994).

Of some 35 grants Javits has awarded through 1996, five both draw on the theory of multiple intelligences ("MI") (Gardner, 1983) and serve economically disadvantaged and minority youth. I undertook a study of three such efforts initially to shed light on the question: How is the theory of multiple intelligences being used to identify poor and minority elementary students for gifted education? Data collection for this initial question opened up additional areas of inquiry: Why were sites that were explicitly committed to using MI to identify underrepresented youngsters for gifted programs drawing on the theory in a very limited way? To understand this, I explored several features of the context in which the assessment efforts developed. Then, in a more speculative vein I envisioned how the currently constrained application of MI might be modified given state policy, leadership, local history, and other contextual forces. Finally, I considered implications of these sites for policymakers who are concerned with improving the identification of gifted youngsters.

THE RESEARCH CONTEXT

At least two bodies of research literature are relevant to an investigation of how MI is being used to identify gifted and talented students from poor and minority populations. One pertains to difficulties associated with identifying poor and minority students for gifted education. A second addresses the applications of MI to the identification of gifted and talented youngsters.

DIFFICULTIES IN IDENTIFYING POOR AND MINORITY STUDENTS

Difficulties in identifying poor and minority students are frequently associated with two issues. One concerns disjunctions between current conceptions of giftedness and traditional identification methods (e.g., Coleman & Gallagher, 1995; Ford, 1994; Frasier 1989a, 1989b; Frasier, Garcia, & Passow, 1995; Van Tassel-Baska, Patton, & Prillaman, 1991). The second pertains to the impact of traditional identification methods on poor and minority students (e.g., Ford, 1995; Frasier, 1989b; Frasier, Garcia, &

Passow, 1995; Kitano & Kirby, 1985; Schmidt, 1993; Swisher & Tonemah, 1991; Tonemah, 1991).

1. Disjunctions between current conceptions of giftedness and traditional identification methods.

"Who is to say in whom the gift may be found and, indeed, what the gift may be?"
- Thomas R. McDaniel (1993)

Conceptions of giftedness have important social and policy implications (Cassidy & Hossler, 1992; Frasier & Passow, 1994; Renzulli, 1986; Sternberg & Davidson, 1986). Federal and state definitions are supposed to be the basis for structuring and funding local gifted education programs (Cassidy & Hossler, 1992; Coleman & Gallagher, 1995; Passow & Rudnitski, 1993). Such conceptions also influence how efforts to identify giftedness are undertaken (Frasier & Passow, 1994; Gardner, Kornhaber, & Wake, 1996). If such conceptions are misguided, then "valuable talents may be wasted, and less valuable ones fostered and encouraged" (Sternberg & Davidson, 1986, p. 4).

Defining or conceptualizing giftedness in adults is far less challenging than conceptualizing giftedness in children. Gifted adults are recognizable because they regularly demonstrate high-level performances in a culturally-valued discipline, practice, or "domain" (Bloom, 1985; Frasier, Garcia, & Passow, 1995; Gardner, 1995; Gruber, 1986; Jackson & Butterfield, 1986; Tannenbaum, 1983). However -- Mozart and Midori aside -- gifted elementary-age students very rarely exhibit such behaviors (Bloom, 1985; Feldman, 1986; Jackson & Butterfield, 1986; Winner, 1996). In fact, many, if not most adults who ultimately become gifted, do not manifest such precocity during their elementary years (see Bloom, 1985; Jackson & Butterfield, 1986). Therefore, giftedness in elementary students must be conceptualized differently than it is for adults and identified on other bases (Bloom, 1985; Jackson & Butterfield, 1986; Frasier, Garcia, & Passow, 1995; Tannenbaum, 1983).

For most of this century, the question of how to conceptualize and identify giftedness in elementary age children was largely answered by IQ testing (e.g., Tannenbaum, 1983; Treffinger & Renzulli, 1986). For instance, Terman (1925) argued that giftedness consists of "the top 1% of ability level in general intellectual ability as measured by the Stanford-Binet Intelligence Scale or a comparable instrument" (Terman, 1925, p. 43). Another common notion is that youngsters who score in the top 3 to 5 percent of intelligence or achievement tests are gifted or should participate in classes for the gifted (Gagné, Bélanger, & Motard, 1993; Vernon, Adamson, & Vernon, 1977).

However, critics have taken aim at the *logic* of IQ-based conceptions of giftedness (Borland, 1986; Ceci, 1990; Sternberg & Wagner, 1993). Their arguments are partly based on the fact that though intelligence tests do a good job predicting success in school (Jensen, 1980; Morris, 1977; Renzulli, 1986; Sternberg & Wagner, 1993), the tests are only weak predictors of adult success in a particular domain (Borland, 1986; Ceci, 1990; Gifford, 1989; Hartigan & Wigdor, 1989; Jensen, 1980; Sternberg & Wagner, 1993). Terman's longitudinal studies of some 1000 "geniuses" -- so-called on the basis of an IQ of 140 or more -- illustrate this point: Few in his sample achieved national or international eminence (Ceci, 1990; Tannenbaum, 1983). Thus, as Borland (1986) has pointed out, conceptualizing and measuring childhood giftedness in terms of IQ contradicts a common justification for offering gifted education: namely, to provide the nation with outstanding adult talent (See e.g., Gallagher & Weiss, 1979; U.S. Congress, 1994; Mitchell, 1994; U.S. Department of Education, 1993).

Nearly all contemporary researchers and practitioners concerned with gifted education now assert that a number of characteristics not measured by IQ tests are important both to conceptions of giftedness and to actual adult success (Maker, 1993; Renzulli, 1978; 1986; Sternberg, 1986, 1988; Sternberg & Davidson, 1986; Sternberg & Wagner, 1993; Tannenbaum, 1983; Torrance, 1978). For example, perhaps the most prominent notion of gifted children at this time is Renzulli's "three-ring conception." It is based on factors extracted from his studies of the qualities of gifted adults who have made important contributions to our culture (Renzulli, 1978, 1986). According to this conception, gifted youngsters, like their adult counterparts, exhibit three, equally important clusters of traits: above average intelligence, creativity, and task commitment.

Another influential contemporary conception is that of Robert Sternberg (Cassidy & Hossler, 1992; Ford, 1995; Frasier, 1989a; Passow & Rudnitski, 1993). Sternberg's triarchic conception of intellectual giftedness is an outgrowth of his triarchic theory of intelligence (Sternberg, 1985). In this view, giftedness arises out of the individual's information-processing capacities; the amount of experience an individual has with a particular task or problem, and his or her ability to function in "real world environments" (Sternberg, 1986a, p. 235). Individuals differ with regard to their strengths in each of these three areas. Furthermore, given that real world contexts differ, Sternberg (1986a) asserts that what is considered intelligent or gifted will vary across contexts and cultures.

A third conception of giftedness that is now exerting influence is Howard Gardner's (Borland, 1986; Cassidy & Hossler, 1992; Ford, 1995; Frasier, 1989a; Passow & Rudnitski, 1993; Schmidt, 1993). Like Sternberg, Gardner's view of giftedness grows out of his theory of intelligence. The most recent version of Gardner's theory of multiple intelligences asserts that each individual possesses at least eight relatively autonomous "intelligences": linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, interpersonal, intrapersonal, and naturalist (the ability to draw on aspects of the natural world to solve problems or fashion products) (Gardner, in press).

Intelligences are "psychobiological potentials" which are available to all unimpaired human beings at birth to process different kinds of information (Gardner, 1983, 1995; Walters & Gardner, 1986). Over time, children's intelligences develop to process and to produce the forms of information (or "symbol systems") available in their environment. Ultimately, individuals are able to draw on various combinations of intelligences "to solve problems or to create products that are valued within one or more cultural settings" (Gardner, 1985, p. x).

According to Gardner, culturally valued products and problem-solving occur within "domains." These are any activities "in which individuals participate on more than a casual basis, and in which degrees of expertise can be identified and nurtured" (Gardner, 1995, p. 202). For example, in American culture, car repair, marketing, robotics, ballet, rap, geometry, and journalism are all domains. It is in efforts that employ the media and materials of such domains that diverse intelligences are developed and meaningfully assessed. In contrast, traditional testing "engages primarily the linguistic and logical-mathematical faculties" as used in school (Gardner, 1991a, p. 85).

According to Gardner, a gifted youngster is one who advances rapidly through a domain of knowledge, due to strength(s) in her intelligences and to opportunities in the environment to develop them (Gardner, 1993a).

Just as scholars' and theorists conceptions have extended beyond IQ-based notions of giftedness, so have policymakers'. In 1972, the federal government adopted the definition advocated by Education Commissioner Sidney Marland following his extensive report on gifted education (Marland, 1971/1972; U.S. Department of Education, 1993). The Marland Report defines gifted and talented children as "those identified by professionally qualified persons, who by virtue of outstanding abilities are capable of high performance" or who show "potential ability" in one or more of the following areas (1) general intellectual ability, (2) specific academic aptitude, (3) creative or productive thinking, (4) leadership ability, (5) visual and performing arts, and (6) psychomotor ability (Marland, 1971, pp. I-3-4). Marland's definition included the notion that gifted and talented students "require differentiated educational programs and/or services beyond those normally provided by the regular school program in order to realize their contribution to self and society" (Marland, 1971, p. I-3).

Over the last 15 years or more, the Marland definition and later modifications to it have grown increasingly evident in state definitions (Cassidy & Hossler, 1992; Coleman & Gallagher, 1995; Ford, 1995; Gallagher & Courtright, 1986; Passow & Rudnitski, 1993). For example, in a recent survey of the departments of education in all 50 states and the District of Columbia (Coleman & Gallagher, 1995), 41 states included the idea of potential giftedness in their definitions and all states had multiple types of giftedness included, rather than just measured cognitive ability.

In 1988, along with the legislation establishing the Javits Program, the federal government again revised its definition to encompass the notion that giftedness was not only manifested or potentially manifested in diverse human endeavors, but that it also crossed cultural and economic lines. According to this latest federal definition, gifted youth are those "with outstanding talent [who] perform or show the potential for performing at remarkably high levels of accomplishment when compared with others of their age, experience, or environment ... in intellectual, creative, and/or artistic areas, [or] possess an unusual leadership capacity, or excel in specific academic fields." Furthermore, "Outstanding talents are present in children and youth from all cultural groups, across all economic strata...." (U.S. Department of Education, 1993, p. 26).

Thus, while some observers assert that gifted education lacks a clear conception of giftedness (e.g., Ford, 1995; Frasier, Garcia, & Passow, 1995; Harris & Ford, 1991), an unfolding notion from both educational policy and theory is that giftedness is a multifaceted quality, potentially manifested in a range of domains by people of diverse cultural and economic backgrounds. Despite this emerging consensus, and despite federal and state policymakers' increasing concern about underrepresented groups (Coleman & Gallagher, 1995; U.S. Department of Education, 1993), broadened conceptions of giftedness are inadequately reflected in local districts' and schools' identification practices (Coleman & Gallagher, 1995; Frasier, Garcia, & Passow, 1995; Tonemah, 1991; U.S. Department of Education, 1993; Van Tassel-Baska, Patton, & Prillaman, 1991).

Several facts on the policy front help to explain this gap between broadened conceptions of giftedness and local identification practices. First, though the states' definitions are not solely based on IQ, the states do describe giftedness partly (and usually first) in terms of intellectual and academic achievement (See Cassidy & Hossler, 1992; Coleman & Gallagher, 1995; Passow & Rudnitski, 1993; Van Tassel-Baska, Patton, &

Prillaman, 1991). Second, several states (including North Carolina and Arizona which are home to two of the assessment efforts in this study), link state funding to high performances on intellectual and academic achievement. Third, while most states recommend using a wide variety of information to identify youngsters, few states mandate these practices (Coleman & Gallagher, 1995). Fourth, though 34 states mandate that gifted students be identified (Coleman & Gallagher, 1995; Passow & Rudnitski, 1993), most of the federal and state parameters of giftedness are not readily measurable (Ford, 1995). Lacking clear measures for most parameters, districts and schools continue to rely on existing approaches, especially standardized intelligence and achievement tests (Borland, 1989; Sternberg, 1986b, Tyler-Wood & Carri, 1993).

Alongside policy, problems associated with theory and measurement contribute to the gap between broadened definitions and actual identification practices. The newer conceptions of giftedness are not easily translated into clear and more equitable measures. For example, as Renzulli himself notes, the evaluation of creativity (one of his three rings) is fraught with difficulty. Creativity tests are seen as biased (Shore, Cornell, Robinson, & Ward, 1991) and lacking construct and predictive validity (Gardner, 1991a; Renzulli, 1986). Adequate assessments and criteria for identifying creativity in young people are yet to be developed (Renzulli, 1986). A second ring, task commitment, is infrequently exhibited before adolescence (Renzulli, 1986), even by individuals who were later recognized as gifted adults (Bloom, 1985). This makes it difficult to identify elementary students using Renzulli's theory, even though early identification is considered crucial for poor and minority gifted youth (Gallagher, 1994; Hartley, 1991; Kitano cited in Smutny, 1996).

Translating Sternberg's theory into clear assessments that might foster equity also poses challenges. Sternberg's suggested assessments are largely paper-and-pencil activities which present relatively novel kinds of problems (Sternberg, 1986a, 1988). Though efficient, these may not capture giftedness as manifested across the domains now acknowledged by federal and most state guidelines or as manifested across the real-world contexts that Sternberg himself asserts matter most (Sternberg & Wagner, 1993). There is also evidence that youngsters do not demonstrate their best thinking in such decontextualized tasks -- that is, in tasks lacking connection to everyday and meaningful activities (Ceci, 1990; Lave & Rogoff, 1984).

In contrast to Sternberg, Gardner and his colleagues contend that children should be assessed by observing them on many occasions over time as they are engaged in domain-relevant tasks. Furthermore, children should not be assessed primarily through paper-and-pencil or verbal measures. Instead, they should be allowed to demonstrate their abilities using more "intelligence-fair" media and materials. For example, to assess students' spatial ability, children could be asked to design buildings using blocks; to assess their musical abilities, they could be ask to make up a tune or sing a song (Gardner, 1991a; Krechevsky, 1991, 1994).

Such contextualized, engaging, and sustained assessments, Gardner argues, are much more likely to reveal the range of students' abilities and provide useful information for advising and placement (Gardner, 1991a). However, in contrast to Sternberg's methods, Gardner's approach is clearly labor and time intensive. It is also likely to require a fair amount of training and practice to use competently (Krechevsky, 1994). Nevertheless, Patricia O'Connell Ross, the director of the Javits Program, asserts

Gardner's ideas have had the most influence on new efforts to identify underrepresented youngsters (Schmidt, 1993).

Part of the aim of this investigation is to detail the tasks and procedures that are being used to make MI feasible for mass identification purposes. If such adaptations prove sound -- that is, if it is reasonable to make inferences about students' abilities from such identification efforts -- then this information could help to close the gap between theory and policy, on the one hand, and practice on the other. If the current adaptations of MI to gifted identification are not sound, then it is crucial to detail their strengths and weaknesses. This information could then enable educators and policymakers to develop stronger, more justifiable assessments. As Borland has asserted, "Until better measures come along in fulfillment of promises made by Gardner (1983), Sternberg (1984) and others, they [IQ tests] will remain among the most useful instruments available to us" (Borland, 1989, p. 113).

2. The impact of traditional identification measures

While there are a variety of approaches to identifying youngsters for gifted education, four practices now predominate. As detailed below, each of these poses problems for the identification of poor and minority students.

Teacher referrals are usually the starting point for identifying students for gifted programs (Borland, 1989; Ford, 1994, 1995; Frasier, Garcia, & Passow, 1995). Teacher referrals would appear to make sense since teachers have sustained opportunities to observe the abilities of their students (Borland, 1989; Frasier, 1980; Roedell, Jackson, & Robinson, 1980). Yet, research provides conflicting signals about teachers' ability to refer students accurately (Adams & Callahan, 1994; Copenhaver & McIntyre, 1992; Ford, 1995; Gagné, 1994; Renzulli, Hartman, & Callahan, 1980).

Some scholars assert teachers do a poor job, tending to identify capable, polite students over less compliant youngsters with greater potential (Adams & Callahan, 1994; Pegnato & Birch, 1959). Others have found teachers' accuracy can be improved through coursework in gifted education and with observer checklists that structure teachers' assessments of students' potential (Copenhaver & McIntyre, 1992; Renzulli, Hartman, & Callahan, 1980).

Whether or not teachers can accurately refer students, it is clear that teachers refer disproportionately fewer African American, American Indian, and Hispanic youngsters (Davis & Rimm, 1989; Ford, 1994, 1995; Frasier & Passow, 1994; Frasier, Garcia, & Passow, 1995; Harris & Ford, 1991). Explanations for this vary: Some assert that teachers have little familiarity with gifted education generally (Copenhaver & McIntyre; Ford, 1994) and have even less knowledge of behaviors associated with giftedness in children from diverse cultures (Adams & Callahan, 1994; Ford, 1995; Frasier, 1989a; Torrance, 1978). Low expectations of minority students are also blamed for lower referral rates of poor and minority students (Ford, 1995; High & Udall, 1983; Kolb & Jussim, 1994). Accurate referrals may also be undermined by the fact that students from some minority groups tend to obscure rather than display their effort and ability in school in order to maintain peer relationships (Fordham & Ogbu, 1986; Garrison, 1989; Mickelson, 1990).

Student grades are frequently used in the identification process. Problems of differential identification associated with using student grades parallel those associated with teacher referrals. Some investigators maintain that teachers hold different expectations for majority and non-majority students, which may affect their grading and instruction (High & Udall, 1983; Howard & Hammond, 1985). Students who especially value group identify, among these many American Indian and African American students, may consciously avoid achieving high grades for fear that this may isolate them from their peers (Ford, 1994, Fordham & Ogbu, 1986; Garrison, 1989; Mickelson, 1990). Given such issues, grades of students from some minority groups may not reflect their actual or potential abilities.

Achievement tests are also widely used in the identification of students for gifted programs (Coleman & Gallagher, 1995; Shore, et al., 1991; Van Tassel-Baska, Patton, & Prillaman, 1991). These tests are logically supported by the notion that future achievement or success in school is predictable from past and current achievement (Shore, et al., 1991). However, achievement test scores also rely on children's prior learning experiences and opportunities (Mercer & Lewis, 1978; Shore, et al, 1991). Given that these experiences and opportunities vary across race and economic lines (Ford, 1994; Heath, 1983; Kozol, 1991; Natriello, McDill, & Pallas, 1990; Ogbu, 1978; Tonemah, 1991), it is not surprising that achievement scores vary along similar lines. On average, middle class white students achieve higher achievement test scores than students from most other groups (e.g., Mullis, Campbell, & Farstrup, 1993; Mullis, Dossey, Owen, & Phillips, 1993). Thus, achievement tests support the identification of disproportionately fewer minority and poor students.

IQ tests remain the central instrument for identifying students for gifted programs (Harris & Ford, 1991; Sternberg, 1986b; Tannenbaum, 1983; Tyler-Wood & Carri, 1993).

IQ tests are said to have many strengths including reliability, validity for school achievement, and objectivity when compared, for instance, with teacher judgments (Borland, 1986, 1989; Kaufman & Harrison, 1986; Robinson & Chamrad, 1986; Shore, et al., 1991; Tannenbaum, 1983; Vernon, Adamson, & Vernon, 1977). Given this, many scholars have argued that IQ tests can be used alongside other identification procedures, especially if tests are selected with care and are properly used and interpreted (Baska, 1986; Borland, 1986, 1989; Kaufman & Harrison, 1986; Shore, et al., 1991; Tyler-Wood & Carri, 1993).

Unfortunately, in efforts to identify gifted students, the tests are often used improperly. Test advocates assert that a group-administered IQ test can be used *to screen* large groups of youngsters and create a smaller, more manageable pool of students from which *to select* youngsters via in-depth individual IQ testing and other identification procedures. Yet, because they are inexpensive and efficient to administer, group IQ tests are often used to select rather than screen youngsters, though group tests are too crude a measure for that purpose (Borland, 1989; Shore, et al., 1991).

Even when IQ tests are properly used, they can contribute to the under-representation of poor and minority youngsters in gifted education (Ford, 1994; Harris & Ford, 1991; Kitano & Kirby, 1985; Schmidt, 1993; Tyler-Wood & Carri, 1993). One key problem lies in the well-documented fact that average IQ test scores differ across groups (e.g., Jensen, 1980; Herrnstein & Murray, 1994; Ogbu, 1978). To illustrate, an IQ score of about 130 is quite commonly used for identification purposes (Gagné, Bélanger, & Motard, 1993). This score falls two standard deviations above the average white IQ, but three standard deviations above African Americans' average. Thus, approximately 2.4

percent of white students meet or exceed this criterion, but only .13 percent of African American students do. Given the normal distribution of IQ scores and the differences in average scores across groups, whatever IQ criterion is established as "gifted," disproportionately fewer African American, Hispanic, and American Indian students will be identified.

In sum, the most frequently used identification methods contribute to the underrepresentation of poor and minority youngsters in gifted education. For gifted education to become more equitable, new identification methods need to be developed and deployed.

EFFORTS TO IDENTIFY GIFTED YOUNGSTERS THAT DRAW ON MI

There are at least two reasons why MI exerts influence on efforts to identify students for gifted education. First, as the above discussion of definitions indicates, MI resonates with the broadened conceptions of giftedness now advocated by scholars and policymakers (e.g., Ford, 1995; Passow & Rudnitski, 1993; U.S. Department of Education, 1993). In both the theory and these definitions, the areas in which human beings may excel extends beyond traditional cognitive and academic realms to encompass the range of human endeavors valued in a society.

Second, the adoption of MI to identify giftedness can also be seen in the context of the larger, "authentic assessment" movement (Plucker, Callahan, & Tomchin, 1996). There is a growing interest by scholars, schools, districts, and states to develop alternatives to traditional, standardized, paper-and-pencil tests (Gardner, 1991a; Madaus & Kellaghan, 1993; Wiggins, 1993a; Wolf, LeMahieu, & Eresh, 1992; Worthen, 1993). Authentic assessments include such approaches as student-generated assessments and reflections, performances in front of peers and teachers, and portfolios of student work. They can entail an examination of students' products as well as students' processes, through vehicles such journals or logs (Stiggins, 1994; Wolf, Bixby, Glenn, & Gardner, 1991; Worthen, 1993).

The calls for more authentic assessments are based in part on arguments about human cognition. According to such arguments, human knowledge and skill are "situated": that is, they are manifested in particular activities, contexts, and cultures (Brown, Collins, & Duguid, 1989; Lave & Rogoff, 1984; Resnick, 1987, 1991; Resnick, Levine, & Teasley, 1991; Rogoff, 1990). Following from this situated view are arguments that traditional test situations -- which are devoid of conversation, computers, books, and other problem-solving resources -- provide very limited insights into what youngsters know and can do (Ceci, 1990; Gardner, 1991; Resnick, 1987; Wiggins, 1989, 1993a, 1993b). Advocates of authentic assessments assert that it is necessary to assess students with engaging problems and a range of problem-solving resources to ascertain individuals' knowledge and abilities (Gardner, 1991a; Stiggins, 1994; Wiggins, 1993a, 1993b; Wolf, Bixby, Glenn, & Gardner, 1991). Such assessments tend to be more "intelligence-fair." They allow youngsters to draw on a range of media and materials, rather than represent their abilities exclusively in language and notations.

The argument for authentic assessment also rests on educational grounds: Since the format of assessment or testing influences classroom curriculum and pedagogy -- teachers teach to the test -- reformers hope that authentic assessments will ultimately yield more engaging learning environments for students (Madaus & Kellaghan, 1993; Wiggins, 1989, 1993a, 1993b; Wolf, LeMahieu, & Eresh, 1992; Worthen, 1993).[1]

Though MI resonates with current conceptions of giftedness and trends in assessment, and though the theory has become enormously popular among educators (Gardner, 1995; Knox, 1995; Levin, 1994; Plucker, Callahan, & Tomchin, 1996), it is difficult to know how and where the theory is being used to identify youngsters for gifted education. As with other educational applications of MI, it is possible that such work is carried out within individual schools or districts and goes unreported by them (Kornhaber & Krechevsky, 1995). When this dissertation began, extensive data base searches revealed four programmatic efforts to use MI to identify gifted youngsters. All of these are associated with the Javits Program (U.S. Department of Education, 1994; Rogers, personal communication, 1995; Ross, personal communication, 1995).[2]

Three of the Javits programs appear to have drawn in some measure on Project Spectrum. Spectrum was a 9-year research project organized by Gardner and David Feldman in 1984. Among other goals, Spectrum sought to discover whether it was possible to identify the relative strengths of the intelligences in young children. Spectrum researchers reported some success in this effort (Gardner & Hatch, 1989; Krechevsky, 1991, 1994). The Spectrum approach to identifying strengths was to fuse curriculum and assessment within the regular classroom. The argument for this classroom-based approach was akin to those made by advocates of authentic assessments: In order to uncover children's strengths, children need experience with engaging problems and materials. Using such problems and materials, the researchers developed a prekindergarten - first grade curriculum involving seven domains (e.g., science, music, mathematics, visual arts, storytelling). To observe and assess more systematically, they devised an accompanying battery of 15 one-on-one assessment activities (Gardner, 1991a; Gardner & Hatch, 1989; Krechevsky, 1991, 1994).

Though Spectrum was not developed to identify gifted youth from under-represented populations, its 15 assessment activities can be used for "selective assessment" (Krechevsky, 1994, p. 6). That is, pieces of the battery might be used by a teacher to evaluate whether a child has strengths in a particular area or domain. For example, teachers who have not noted any particular talent in writing or mathematics among some of their students, have used Spectrum materials, such as a model of their own classroom and figurines of the students in it, to see if these students show strengths in understanding social interactions. By giving youngsters a variety of small machines to put together and take apart, teachers have uncovered unusual ability in spatial relationships and bodily-kinesthetic skills.

Among the Javits Programs that have drawn on Spectrum is the Javits 7+ Program in Community School District 18 in Brooklyn, New York. Javits 7+ uses classroom-based curriculum and assessments derived from Spectrum activities to identify and nurture children's strengths in the early elementary grades (Baldwin, 1994; Metis Associates, 1994). During the fall, children have extensive classroom experiences in activities drawing on each of the seven intelligences. Then during the assessment phase in December, students are given open-ended assignments for each of the activities, which are carefully observed against a number of criteria. Through these approaches Javits 7+ seeks to identify students "at promise." These youngsters then receive enriched curriculum to enhance their "prospects for admission into the district's existing gifted program" (Metis, 1994, p. 1).

Another Javits program that made use of Spectrum approaches is Montgomery County Maryland's Early Childhood Model Gifted Program. The Model Program sought to "demonstrate the effectiveness of Howard Gardner's concept of multiple intelligences ... as a vehicle for identifying and nurturing underserved and culturally diverse gifted students" (U.S. Department of Education, 1994, p. 25). Like Spectrum, it provided assessments in familiar, domain-based classroom activities.

DISCOVER III, directed by Professor C. June Maker of the University of Arizona, is a Javits-funded effort which draws on MI but does not draw on Spectrum (Maker, 1992; Schmidt, 1993). It is established in nine local education agencies in Arizona (U.S. Department of Education, 1994). According to Maker, MI provided DISCOVER with a conceptual framework for viewing intelligence in cultural context and in terms of problem solving ability (Maker, 1992). She and her colleagues have devised a diverse set of five assessment activities, some of which entail hands-on tasks. Unlike Spectrum, these are not intended to be part of the regular classroom environment.

Finally, the Charlotte-Mecklenburg Schools developed two identification procedures that draw on both DISCOVER and Spectrum. These two assessments are not extensively embedded in classroom environments. However, pre-assessment lessons, which use activities similar to those administered during the assessment, are taught in the weeks preceding the actual assessment. Like Javits 7+ and Montgomery, Charlotte-Mecklenburg's Project S.T.A.R.T. sought to identify kindergarten and first grade students from poor and minority backgrounds for enrichment classes. These classes are meant to increase the chances of identifying traditionally underserved youngsters in the district-wide assessment for gifted education (Charlotte-Mecklenburg Schools, 1994a; U.S.

Department of Education, 1994). Along with the S.T.A.R.T. assessment, Charlotte-Mecklenburg developed the Problem-Solving Assessment (or "PSA"). The PSA is used beginning with second graders to identify them for gifted program services, which begin in the third grade. Unlike S.T.A.R.T., the PSA was not funded by Javits. However, it evolved in part with the expertise of the S.T.A.R.T. staff and drew on some similar methods. The PSA has displaced the earlier method of traditional IQ and achievement tests to become the predominant means for identifying youngsters in the county.

A few papers examining one or another of the three programs I am investigating have been published or are in preparation. Each of these focuses on the programs' outcomes or on statistical characteristics of the assessment instruments.

For example, Adams and Callahan (1994) have statistically analyzed an MI checklist produced by the staff of Montgomery County's Model Program. The checklist was used by teachers to document students' abilities in different intelligences as manifested in classroom performances. The researchers found that teachers' intrarater reliability on the checklist was moderately high (Adams & Callahan, 1994; but see Chapter 4).

Various aspects of DISCOVER's reliability have been carried out by Maker's graduate students. For example, two studies of observer judgments by Giffiths (n.d.), "suggest that inter-observer reliability has been obtained for the DISCOVER assessment process" (Griffiths, n.d., p. 2; but see Chapter 2). In a comparison of the Raven's Progressive Matrices and DISCOVER III, Romanoff (n.d.) has found that DISCOVER III is more consistent in identifying youngsters over a four-year period.

An investigation by Reid, Udall, Romanoff, and Algozzine (in press) has revealed positive correlations among tasks for different intelligences (in contrast to claims by Gardner that the intelligences are relative autonomous), and between the PSA and the Matrix Analogies Test, a more traditional standardized measure. They have also found that the PSA identifies a more diverse group of youngsters than the MAT would have.

The difficulty in interpreting such findings is that the studies in which they are reported reveal little about the nature of the tasks, the procedures used to collect and document data from students, or the methods of evaluating students' performances that lead to actual identification. Even if these studies are stellar from a statistical vantage point, their findings are supportable only if they are based on assessments whose tasks, procedures, and methods of evaluation are themselves adequate. Examining these assessments' tasks, procedures, and evaluation methods is fundamental to this dissertation. Therefore, it will provide information needed to interpret existing and future studies of these assessment efforts. In addition, it should shed light on the strengths of the assessments, illuminate areas which may be improved and, I hope, ultimately foster more equitable alternatives to existing identification approaches.

RESEARCH DESIGN AND METHODS

THE RESEARCH QUESTION

The initial question driving this dissertation is: How is MI theory being used to identify poor and minority elementary students for gifted education? To explore this question I describe the following assessment components: the activities and tasks that are used; the procedures for administering these activities; the procedures for documenting students efforts; and the means by which information gathered about students is evaluated. I then analyze whether claimed increases in the proportion of identified students from poor and minority populations can reasonably be associated with these assessments and with MI theory.

DATA COLLECTION

The data for this study was gathered from three sites that associated increases in the identification of poor and minority elementary students for gifted education with MI-influenced assessments. Each of the three have also evolved in some measure out of the federal Javits Program. These three are DISCOVER, based at the University of Arizona at Tucson, Charlotte-Mecklenburg Schools' Problem Solving Assessment, and Montgomery County's Early Childhood Model Gifted Program. A fourth site, Javits 7+ in New York City's District 18, was in the midst of personnel changes when this research began and could not grant me access.

To describe and analyze these assessment efforts, I have collected qualitative data from observations, interviews, and documents:

<u>Observations</u>

I spent four days in each of the three sites. To help me see how different settings might alter the identification practice, I visited two schools per site. During site visits in Arizona, I was a participant observer in the administration of the assessment to children. I was also a participant observer in meetings of assessors as they considered and evaluated children's performances using the assessments. In Charlotte, I was solely an observer of the administration of the assessment. I was also largely an observer of meetings in which assessors evaluated children's performances on the assessment.

Occasionally, though, I did pose questions to the assessors about their evaluation methods. In Arizona and Charlotte, I audiotaped and transcribed assessors' discussions as they evaluated children's performances on the tasks. In Montgomery County, children are not given tasks or assessment materials. Instead, they are exposed to a variety of materials in the classroom, and teachers are asked to record information about the students on the MI Checklist twice a year. In Montgomery County, I was able to observe four classrooms to get some understanding of the materials and activities upon which teachers' observations are based.[3] This understanding was expanded during interviews with teachers. I recorded information from my observations at the sites in fieldnotes, supplemented by photographs and video.

Interviews

For each site I conducted, audiotaped, and later transcribed seven or eight semi-structured individual interviews. The interviews were with individuals who helped to design the assessments, with individuals who participate in the process of evaluating and identifying students for gifted services, and with the principal and/or educator responsible for gifted services in each of the schools. (See Appendix A: Interviewees.)

Most interviews were conducted by phone during a six month period following the site visits (these took place in October and December 1995). Most of the interviews for DISCOVER and Charlotte lasted 1.5 to 2 hours. The shortest lasted a half-hour with one school principal. The longest was 3.5 hours. In Montgomery County, most of the interviews lasted one to two hours. Two interviews with classroom teachers in Montgomery County lasted a half-hour. One was three hours. Additional information was gathered from most interviewees via mail, phone calls, and other electronic media through April 1997.

Interviewees discussed previous approaches to identifying gifted children used by the sites, and the MI-influenced assessment materials and procedures now in use. They also discussed administration, documentation, and evaluation procedures, as well as how they were trained for these. In addition, they were asked about outcomes they associated with using the new assessments. (See Appendix B: Interview Guide.)

Interviewees were asked if they wished to remain anonymous either for the duration of the interview or during specific portions of it. All gave permission to use their names. In three or four instances, people did not want particular remarks recorded or credited to them. These wishes were followed.

Along with formal interviews, I had opportunities to sit in on several meetings where the MI-influenced assessments were discussed and to converse with teachers, school principals, district administrators, assessment designers, and others involved with the assessments. I have also tape recorded and transcribed these meetings.

Documents

In addition to observations and interviews, I gathered a variety of documentary data from the sites. These included grant proposals, observer training manuals, observer instructions, and various checklists used to record information about students.

DATA ANALYSIS

To describe the materials, administration and documentation procedures, evaluation methods, and outcomes I began coding fieldnotes and transcripts from interviews, observations, and meetings into four large categories: "Tasks" included all descriptions of activities children participated in and from which information about them was gathered. "Procedures" included data about what teachers or observers did to instruct or guide youngsters in using materials from which information about students was gathered. "Instruments" included information about the documentation materials used by teachers or observers to record students' activities or performances. "Evaluation" was applied to data describing how adults identified students based on information they had collected. "Outcomes" was used to code data about who was identified under the new assessment system. It was also applied to other changes that interviewees associated with the use of the new methods. Each of these eventually yielded several subcategories.

In addition, I applied the code "reliability" to information about inter- or intra-rater reliability. Another code, "reliability of students' performance," was applied to information about students' test-retest reliability. The code "validity" was applied to information indicating that evaluations of students' ability made from the MI-influenced assessments conformed with other assessments of students, for example, their grades, teacher evaluations, classroom performances, or products and performances created outside of school. (See Appendix C: Coding Scheme.)

The issue of whether these or other assessments are valid is increasingly complex. To establish validity requires constructing an argument from a variety of evidence that supports the use of the assessment for a particular purpose (Cronbach, 1989; Messick, 1989; Shepard, 1993; Wiggins, 1993b). At least that is the current approach to establishing validity within the realm of traditional assessment. For authentic or alternative assessments, there is little agreement about what and whether technical standards of reliability and validity should be applied (Worthen, 1993). Thus, the application of the "validity" code in this study is only a beginning effort. Given the preceding discussion of grades, teacher evaluations, achievement and IQ tests, it is unlikely that concurrent validity for these assessments can be gleaned primarily by comparing them with traditional measures. Validation will require more extensive, and likely longitudinal, investigations of students' performances in and out of school.

To analyze whether it is reasonable for each site to associate increases in poor and minority students' identification for gifted education with its new assessment and with MI, I established two sets of conditions. The first set includes five "general" conditions, each of which is necessary to make reasonable inferences about students' abilities from *any* assessment (Cronbach, 1990; Sattler, 1992). This set must be in place to associate claimed outcomes with the assessments in question: (1) Care is taken to ensure children understand the tasks; (2) Care is taken to ensure children do their best work on the tasks. Without these first two conditions, it is impossible to know whether children's performances on the assessment represent their abilities. (3) Assessors must have training to administer and evaluate the assessment. Given that these assessments are not paper-and-pencil tasks scored by a machine, assessors should have training commensurate with the demands of the assessment process. (4) There are clear procedures for scoring student performances. That is, the bases for scoring students should be clearly articulated and used in practice. (5) Assessors' judgments are reliable. It is important to know whether similar student performances are judged similarly.

The second set of conditions includes three "MI-specific" practices. These are needed to associate the assessment with MI theory: (1) The assessments should be broadened beyond the traditionally tested linguistic, mathematical, and spatial abilities.

This condition reflects a central tenet of MI: that all people possess several abilities beyond those that are typically tested (Gardner, 1983). (2) The assessments must be "intelligence-fair" (Gardner, 1991a). That is, they should allow students to demonstrate strengths and to be identified using media appropriate to spatial, bodily-kinesthetic, musical, and other intelligences. (3) The assessments should be "domain-based." That is, they need to allow students to be identified based on performances in cultural practices or domains. This aligns with Gardner's notion, discussed earlier, that intelligence entails an ability to make products or solve problems valued in one or more cultures. For example, an assessment of linguistic ability should not focus on antonyms and synonyms. Instead, such an assessment might ask children to write a story or describe an object -- linguistic abilities valued in the wider culture. Such an assessment can draw on culturally valued criteria for evaluation, such as the presence and quality of the plot, characters, and description.

Each of these eight conditions were coded in a trivalent way. This enabled me to identify information that supported the condition, undermined the condition, or information that was relevant to the condition but neither supportive nor countervaling. (See Appendix C.)

As data collection, transcribing, and coding proceeded, it was clear that a third set of codes was needed to track information about the constraints on implementing the theory. Given this, I evolved a set of codes for "context." These enabled me to highlight features of the environment that influenced the assessment, including links between curriculum and assessment, local history, state policy, organizational setting, resources, and leadership. (See Appendix C.)

To search, code, and sort approximately 1400 single-spaced pages of data in a thorough fashion, I relied on FolioVIEWS Infobase Manager Version 3.1 (Folio Corporation, 1996). Among other things, this computer program rapidly scans data by keywords, codes, and combinations of codes. I then printed coded and surrounding contextual information by category. Thus, to the detriment of forests near and far, I was able to review, reconsider, and analyze data even during those rare moments away from my computer.

BIASES

In carrying out this work, I have in part tried to understand the role that MI theory has played in developing more equitable assessments for gifted education. It is worth noting that my foregrounding of MI may not entirely overlap with the perspective of the assessments' designers. Instead, their foreground may be the desire to increase equity in gifted education, with MI providing a backdrop for this. I nevertheless investigate the MI-specific conditions in order to understand whether MI -- as foreground or background -- is realized in the actual practice of the assessment.

Another potential bias in this work is that my own understanding of MI has been influenced by several years' work at Harvard Project Zero, the organizational locus of Gardner and MI. Despite this, I do not believe I am predisposed to view efforts entailing MI in a more positive light than they might warrant: Project Zero is not dependent on the theory; the research group has had only three of some 24 funded projects built around MI in the last 14 years. Rather, I am motivated to discern, in part, whether others who claim that MI is useful in enhancing equity have methods that can support such assertions.

I am exploring such claims because of my desire to see (or bias toward) publicly-defensible methods of enhancing access to enriched and challenging curriculum of the sort offered in programs for the gifted. I believe these methods should be able to withstand the scrutiny of those who question efforts at equity. Thus, if designers assert that they draw on MI, it is useful for them to be able to substantiate that assertion. Furthermore, if designers or districts want to associate enhanced equity with an assessment in a tenable manner, they need to be able to show how they meet the five general conditions outlined above. It is important to emphasize here that to associate claims of enhanced equity with an assessment, the general conditions need to be met; the conditions associated with MI do not. Those are only needed to associate the assessment with MI.

To the extent that this work reveals weaknesses in the identification efforts, my hope is that this information will provide practitioners, policymakers, and theorists with insights for developing clearer and stronger assessments. By understanding the strengths and promise of these assessments, I hope that the strengths of more poor and minority youngsters may be more fully realized.

1. In the meantime, authentic assessments can create additional obstacles to minority students. Youngsters whose education emphasizes skill and drill instruction may be less likely to know what to do with the more open-ended and hands-on problem solving authentic assessments can call for (Comments by Eva Baker, October 1991, Boston, Massachusetts).

2. A fifth, Project Excel, surfaced after I had completed my site visits. Project Excel was designed by Margie Kitano and Rosa Perez to serve poor and bilingual youngsters in San Diego, California (Smutny, 1996).

3. When access to the site was granted, I had been promised opportunities to observe meetings during which information about students was evaluated for identification purposes. However, over 20 attempts between January and April 1996 to arrange for such observations proved fruitless.

Chapter 2
DISCOVER III: PIONEERS

INTRODUCTION

In this chapter and each of the two that follow, I examine one effort to draw on MI to identify underrepresented youngsters for gifted education. Each of these three chapters first sets the identification effort in its theoretical, historical, and community contexts. Following this, the assessments themselves are described. Finally, I analyze the assessment in light of the eight conditions discussed in Chapter 1. In the concluding chapter, I consider contextual forces that have, and will, shape each of the assessment efforts.

THEORETICAL BASES OF DISCOVER III

DISCOVER III is a broad-scale project to improve gifted education, especially for underserved youth. It entails curriculum and staff development, as well the development of, and research on, its own method of identification. The acronym combines the project's lengthy name: **D**iscovering **I**ntellectual **S**kills and **C**apabilities While Providing **O**pportunities for **V**aried **E**thnic **R**esponses, with the fact that it is the third in a series of efforts to understand variations in problem solving among different cultural groups.

DISCOVER is the work of Professor C. June Maker and her colleagues at the University of Arizona's College of Education in Tucson. The DISCOVER project formally collaborates with nine local schools and districts across Arizona. As a result of publications, lectures, and consulting by team members, the DISCOVER method of identification has been adopted by numerous schools and districts in the United States and in several foreign countries.

Maker acknowledges that the theoretical origins of DISCOVER draw largely on Gardner's work (Maker, 1992, 1993; Maker, Nielson, & Rogers, 1994; U.S. Department of Education, 1994). According to Maker (1992, 1993), MI justifies examining both a range of intellectual strengths and problem solving in activities valued by particular cultures (Maker, 1992, 1993). It "provides a helpful way to examine giftedness across and within cultures because of its inclusion of cultural factors as important influences on the development and expression of abilities" (Maker, 1993, p. 70).

Maker asserts Gardner has also established useful and comprehensive criteria for authentic assessment. Paraphrasing Gardner (1991a), she notes these criteria are:

> [an] emphasis on assessment rather than testing; assessment as simple, natural, and occuring on a reliable schedule; ecological validity; "intelligence-fair" instruments; multiple measures; sensitivity to individual differences, developmental levels, and forms of expertise; use of intrinsically interesting and motivating materials; and application of assessment for the student's benefit (Maker, 1994, p. 20).

Along with Gardner, DISCOVER III assessments draw on the work of Getzels and Csikszentmihalyi. These scholars posited a continuum of problem-solving types which range from closed to open-ended. The former entail problems and methods that are clear to both the presenter and problem solver, and for which there is an existing correct answer. An example is an arithmetic computation problem. On the other end of the continuum are ill-defined problems, whose methods are unknown to either the presenter or the solver, and for which there may be many acceptable solutions. An assignment to develop energy-efficient modes of transportation illustrates this sort of problem. Maker and her student Shirley Schiever later expanded on this continuum, (Maker, 1992, 1993). Members of the DISCOVER team devised assessment activities that span this continuum. (See Figure 2.1.)

Figure 2.1: The Continuum of Problem Types Used in DISCOVER Tasks
(Adapted from Schiever, 1991)

	Problem		Method		Solution	
	Presenter	Solver	Presenter	Solver	Presenter	Solver
Type						
I	Known	Known	Known	Known	Known	Unknown
II	Known	Known	Known	Unknown	Known	Unknown
III	Known	Known	Range	Unknown	Range	Unknown
IV	Known	Known	Unknown	Unknown	Unknown	Unknown
V	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

The problem solving continuum informs Maker's concept of giftedness: "[T]he ability to solve the most complex problems in the most efficient, effective, or economical ways" as well as the ability to solve "simple problems in the most efficient, effective, or economical ways" (Maker, 1993, p. 70).

HISTORICAL BACKGROUND OF THE DISCOVER ASSESSMENTS

Maker's conception of gifted individuals as skilled problem solvers arose from her early research on adults who were successful in their careers, even though they had gone through school with disabilities suffered early in life. A commonality was "the fact that they were all, in many ways, problem solvers.... that in general they looked at problems as challenges to be overcome, rather than something to stop them." This research led Maker to "a concern with the narrowness of definitions of giftedness and the people who were being placed in programs."

Later while teaching at the University of New Mexico, Maker saw:

> [an] even larger group of children who were being overlooked, who were not really disabled but again who had some perceived weakness that was interfering with people viewing them as having a strength in some area. And that was basically Mexican American kids, whose home language was not English, and anybody with a brown face, basically.

With some of her university students, Maker began to develop a different approach to identifying gifted youngsters. In this approach, teachers administered "problem-solving tasks" to a classroom of students, while a graduate student noted particular behaviors for all the students on an observation form that was laid out in a grid. Maker noted that while this initial effort highlighted the use of observation, it also underscored the difficulty in trying to observe whole classrooms.

In 1981, Maker moved to the University of Arizona. Her colleague, David Berliner, referred to her to Frames of Mind shortly after its publication in 1983. The following spring, and every other year thereafter for about the next decade, Maker taught a graduate course in multiple intelligences. Students in this course administered problem sets to an individual age 12 or over who was highly competent in one or more of the seven intelligences posited by Gardner. They then interviewed the individual and analyzed his or her performance. Judith Rogers, who became the coordinator of the DISCOVER III team, was in one of the MI classes. Maker reported that Rogers "kept saying 'we really needed to do these [tasks] for kids under 12' ... [But] below 12 these problem-solving tasks didn't really work that well." Furthermore, these were time intensive, one-on-one investigations.

Several spurs prompted Maker and her colleagues to devise assessments for younger children. In 1991, the Tucson Unified School District requested Maker's help in implementing new methods of identifying underserved youngsters. Since 1984, the district had been monitored by a committee established by the U.S. Office for Civil Rights. This monitoring followed a complaint that minority children were given an inferior education, including limited access to college prep and gifted classes. Eight years later, the monitoring committee concluded that gifted education in that district was "still dominated by white students." The committee "admonished the Gifted and Talented Education ... for sluggish progress in minority recruitment. The report recommended dismantling the program's management" (O'Connell, 1992, p. 1). Instead, the GATE program director was instructed by the district to increase minority participation in the gifted program by the end of that school year (1991-1992) (O'Connell, 1992).

At about this same time, Rogers began an internship in the Tucson Unified School District. There, she and staff from Tucson's Gifted and Talented Education program conducted informal observations of kindergarteners in schools with high minority student populations. Kindergarteners who seemed to possess potential for gifted and talented education were given some new assessment materials that the district hoped would make identification more equitable. The assessment included two non-verbal, spatial problem solving components. One was Raven's Colored Progressive Matrices, a standardized psychometric test of reasoning for young children based on visual patterns.[1] The other entailed tangrams, flat geometric pieces that can be combined to make specified shapes, which were later incorporated into the DISCOVER identification process. From the fall of 1991 into 1993, Maker, Rogers, and other graduate students worked with TUSD to help expand the assessments beyond spatial tasks to include the two other realms recognized by the State of Arizona: quantitative and verbal. According to Aleene Nielson, a member of the DISCOVER team, observations of the children working on these tasks helped the designers to devise a checklist of behaviors associated with strengths in the different intelligences. (The checklist is described below, under "Description of DISCOVER III Tasks and Procedures.")

About the same time that TUSD was seeking help, Maker was approached by Dorothy Sisk, a professor now at Lamar University, to participate in Sisk's new Javits-funded effort, Project STEP-UP. STEP-UP sought to increase identification of youngsters from low income, culturally diverse backgrounds who scored near the gifted level on traditional tests. Sisk's plan was to provide these youngsters with enriched curriculum in classes of 18 students, and have the classes taught by the same teacher for three years. Sisk anticipated that many of these youngsters would then be formally identified as gifted via traditional tests.

Sisk asked Maker to coordinate STEP-UP work in Arizona. However, Maker and Sisk had a basic disagreement: According to Maker, Sisk "believed we needed to find them [underrepresented youngsters] using the same method that we use with other kids." After locating four potential STEP-UP sites on the Navajo reservation, Maker told Sisk:

> [I]f I'm going to be involved in this, then I need to be able to put my own ideas into practice. And since we don't really have a whole lot of ways to identify kids that are appropriate for this population [Navajos], why don't you let us design some? So we did. And that's how we did the first DISCOVER assessments.

Assessments Maker used in STEP-UP, including storytelling, tangrams, a construction task involving Pablo® pieces (described below), were later refined with Javits funds for DISCOVER III.

INQUIRY INTO DISCOVER III ASSESSMENTS

The observational data about DISCOVER III come from assessments of fifth graders in two schools held during the first week of October, 1995. The assessment/observer team was led by Dr. Judith Rogers, who worked with DISCOVER for several years and who contributed to the design of the assessment. Though she is only one of a several possible team leaders, Rogers led many, if not the majority, of DISCOVER III observations on the reservation during the year of my visit.

The two schools in which I observed were among the original four Maker located when she began her STEP-UP work. Maker initially worked with these four schools, because, unlike those closer to Tucson, the schools were able to modify their classroom size to approximately 18 students, in line with Sisk's design. Both schools are within the Navajo Nation, whose population of between 150,000 and 200,000 live in an area the size of West Virginia, that extends across northern Arizona, western New Mexico, and southern Utah.

Chinle Elementary School (CES), is a public school in Chinle, Arizona. The town is most well known for the Canyon de Chelly, a breathtaking canyon at the bottom of which lies a shallow stream and a compact collection of ancient Anasazi ruins. The Canyon attracts busloads of tourists, and so Chinle has the trappings of many other American communities: A Holiday Inn, fast food restaurants, a supermarket, a hospital. Yet, cattle graze unfettered and unfenced at the edges of parking lots, and from the late 1980s until the mid-1990s, the nearest bank was 70 miles away (Bradsher, 1994).

CES is part of the Chinle Unified School District, a seven-school district encompassing 4400 students spread out over 7200 square miles. It is administered by a local school board. At the time of my visit, CES had 752 students in grades 4-6. About 98 percent of the students are Native American, almost all of these Navajo. Almost 85% of the students are on free or reduced lunch. Approximately ten percent come from homes with no running water or electricity.

CES was built about 1991. It is an attractive, single-story tan and tourquoise structure, that is well-funded and well-equipped. There are 50 teachers, including music, art, PE, and reading specialists. All the classrooms and the library have several computers. Teachers are free to use the color copier and to draw on abundant office supplies.

Chinle Boarding School (CBS) is located 13 miles away in Many Farms, Arizona. Many Farms is a much smaller community than Chinle. Its unemployment rate is about 75 percent. The town is dominated by the Boarding School, which is run by the Bureau of Indian Affairs, and by other schools run by the Chinle Unified School District. Beside the school buildings are modest homes, many the residences of BIA teachers. Beyond this, and throughout most of the reservation, are large tracts of open dry grassland, dotted by small houses and traditional homes, hexagonal hogans, each facing east to meet the morning sun, and many surrounded by small stands of corn and other crops.

CBS serves 500 K-8 students, all American Indian, almost all Navajo. According to the principal, nearly all the students are poor. About one quarter of the students board at the school. The school staff reported that most boarders are placed there by social services. Others board because they live too far away from bus routes.

The school building is a single-story tan structure. Inside it is very clean but extremely spartan. A glass display case stood empty in the school's entryway. Along some of its numerous corridors, was a narrow cork strip roughly five feet off the ground onto which some student essays were tacked. Inside the classrooms, computers were scarce. Due to recent federal budget cuts, which drastically reduced BIA funding, gym, art, shop, and other electives had been cut, and teachers in those areas were being reassigned. Dorm staff had also been reduced.

In both schools many youngsters are bilingual, though they may lack mastery in either Navajo or English. Some are ESL, with Navajo being the primary language.

EFFORTS TO IDENTIFY GIFTED YOUNGSTERS PRIOR TO DISCOVER III

Both the Boarding School and Chinle Elementary had difficulty identifying and serving its most able youngsters prior to adopting DISCOVER's assessments. Interview data suggests that cultural factors as well as traditional identification methods depressed identification rates.

On the cultural front, staff at both schools reported that the Navajo find it generally unacceptable to single out anyone. Nor is it appropriate to "'stick out from others'" (Hartley, 1991, p. 58). Thus labelling someone as gifted -- or seeking to be identified as gifted -- violates a cultural norm.[2] To the extent that Navajo youngsters have been recognized for their potential, it was for their physical strength or their ability to master the ceremonies and songs of Navajo culture.

Alongside cultural practices, traditional psychometric tests used in identifying giftedness proved problematic in several respects. The BIA recognizes gifted and talented students as those possessing potential in six areas, akin to those described by Marland (1971/1972): academic achievement, intelligence, critical thinking, creativity, leadership, and psychomotor skills. The BIA guidelines thus potentially allow recognition of talent in cultural domains. For example, creativity or leadership might be demonstrated in areas valued by the Navajo. Yet, two staff members at CBS independently reported that the BIA looks for numbers. As the principal put it, BIA administrators "want scores, and how do you score a student who can sing and dance in their native ways?"

It is hard to know whether children at CBS were actually identified prior to DISCOVER. The principal asserted that in the past students were identified by teacher observation and recommendation. However, two teachers independently noted that if any students had been identified previously, the staff was not notified about it, and the children were not served.

For Chinle Elementary and other schools in the Chinle Unified School District, identification is supposed to follow Arizona's state guidelines. That is, children are supposed to be tested using nationally normed instruments, and those who score at or above the 97th percentile are supposed to be provided with special services for gifted youngsters (Arizona State Department of Education, 1992).

In line with state policy, prior to about 1990, the district administered the Iowa Test. According to Susan Bartley, the director of All Can Excel, Chinle Unified's enrichment program, all students in the district scored below the mean on the Iowa Test. "Every single one of them" including the Anglo youngsters, "so," she added facetiously, "are you telling me it's genetic?"

Some years before the Iowa Test, the district gave the Cognitive Abilities Test, which has verbal, quantitative, and nonverbal/spatial components. It is supposedly better at identifying minority and economically disadvantaged youth (Kaplan & Sacuzzo, 1993).

Using the 97th percentile cut off score that Arizona requires then yielded a cohort of about 15 Anglos and 10 Native Americans in a district that is 98-99 percent Native American.

The adoption of Maker's methods in Chinle Unified was spurred partly by Bartley's attempt to demonstrate to the school board that there actually were gifted youngsters in the district. She administered the Ravens, on which 70 percent of the students scored above the mean. The adoption also spurred by the state's threat to fine the district for not providing gifted education.

Financial concerns also motivated CBS's interest in DISCOVER. The BIA demanded that the boarding school move from teacher observations and recommendations to a more formal system of identifying gifted students. The demand was heeded because the BIA provides extra funds for all students identified as gifted, up to ten percent of the school's population.

IDENTIFICATION OUTCOMES USING DISCOVER III ASSESSMENTS

Using DISCOVER III assessments in CBS, at the time of my visit, 52 youngsters, slightly more than the BIA's allowable 10 percent were identified as gifted. In Chinle Elementary School, approximately one-third of the fifth graders were identified in the three classrooms that DISCOVER assessed during the 1995-1996 school year.[3]

In both CBS and CES, students identified through the DISCOVER process are placed in classrooms with teachers who have had additional training to support gifted learners through enriched curriculum and other means. Because cultural practices proscribe singling out individuals, identified children, as well as those who are not, have access to teachers with additional training.

While identification rates have gone up in both schools, is it reasonable to associate these increases with the DISCOVER identification process? For this question to be answered, it is necessary to look at the assessment tasks, how they are administered, and how information about students' performance on them is evaluated. After these descriptions, I analyze the assessment in terms of the general and specific conditions introduced in Chapter 1.

DESCRIPTION OF DISCOVER III TASKS AND PROCEDURES

The DISCOVER identification process has been tailored to students at different grade levels. I observed and interviewed people primarily on the activities geared for students in grades three through five.

DISCOVER assessments for each grade cluster involves two sets of tasks. One set is fairly traditional, the other less so. The two sets are carried out on different days. Both are administered in the students' usual classroom to help make the assessment comfortable for the youngsters and to make the process "less instrusive."

THE TRADITIONAL TASKS

The more traditional tasks are made up of a short-answer math worksheet and a writing sample (if age appropriate) that DISCOVER has devised. These two tasks are given to students on two separate days, within a few days of each other. Both tasks are untimed (Maker, Rogers, & Nielson, 1995). According to Rogers, the amount of time the task takes is based primarily on the "upon the engagement of the kids."

The Math Worksheet

For students in third through fifth grade, the math sheet consists of a single 8.5 x 11 inch page containing four sets of problems. These problems move from closed to more open-ended problem types. There is no Type V problem (one whose methods and solution are unknown by the presenter and the test-taker) because, as Aleene Nielson put it, "we've been conditioned that math has right answers." The math assessment is administered by the teacher according to a set of written instructions (Maker, Rogers, & Nielson, 1995).

Task 1 consists of nine arithmetic problems. These include two- and three-digit addition and subtraction, one- and two-digit multiplication and division, and one addition problem involving fractions (1/4 + 2/4 =). The teacher instructs the students to "Solve problems 1 through 9 and then put your pencil down so I know that you are ready to continue" (Maker, Rogers, & Nielson, 1995, p. 10).

Task 2 entails three magic squares, each containing 3 rows and 3 columns. One box involves subtraction of two- and three-digit numbers, one involves multiplication of one-digit numbers, and one box is left blank. The directions call for the teachers to demonstrate how to solve magic squares using an example provided in the instructions. The children are then instructed to "Add numbers to the incomplete magic square to create your own problem. Solve the three magic square problems and put your pencil down when you are finished" (Maker, Rogers, & Nielson, 1995, p. 10).

Task 3 contains four sets of numbers. Each consists of three, one- or two-digit numbers followed by a blank line (e.g., 8, 32, 4 _______; 7, 9, 63 _______). For each set, the students must devise correct arithmetic problems. The teacher tells them "You are to use only these numbers and write addition, subtraction, multiplication, or division problems that are correct on the line to the right. I'll know you are finished, when I see your pencil down on your desk" (Maker, Rogers, & Nielson, 1995, p. 10).

Task 4 calls for the students to "Write as many problems as possible that have 18 as the answer." Before the students begin, the teacher demonstrates how such problems can be constructed using 6 as an example. When most students have finished, the written instructions call for the teacher to ask the children to check their work and complete any unfinished problems.

The Writing Task

For the writing task, the teacher is supposed to tell the students to "Write a story about anything you want to write about. You can write about something that happened to you or something you make up or imagine. Make the story as long as you wish, and do not worry about how to spell words. You may write in any language you would like to use. I will not grade you at all. I am only interested in your story" (Maker, Rogers, & Nielson, 1995, p. 8). The students are given paper. They are supposed to be given as much time as they need.

Although the traditional assessments were designed to be administered by classroom teachers, beginning in the fall of 1995, members of the DISCOVER team began administering the writing and math tasks in 12 classrooms that they were studying intensively. The reason for the team-based instructions was, according to Aleene Nielson, "that teachers were not following the instructions as carefully as needed to be done for research purposes." Rogers' interpretation was the same. With team members guiding the task, directions were "more consistent across sites."

Students' Experience of the Traditional Tasks

As the descriptions above indicate, the traditional tasks look quite school like: Directions are read to the class as a whole, children are given paper and pencil to work with, and the completed work is collected at the end for scoring elsewhere. Not surprisingly, children treat the tasks as school-like: They work quietly and on their own, even though there are no explicit instructions to do so: As Rogers put it, "... they get engaged in doing the math, but it's just an individual effort.... The writing is basically the same way. Nobody says they can't talk, but it's like any other writing you're given in a class. So they act more like they're doing school tasks."

Children vary in the degree to which they engage. For example on the writing task, some complete it in ten minutes. On the opposite end of a continuum, one girl at Chinle Elementary School asked to take her writing home so she could add sound effects to it. She brought the completed piece the next day accompanied by a tape of sound effects. In general, however, each of the two traditional tasks is completed in under an hour of class time.

Evaluation/Scoring of the Traditional Tasks

All DISCOVER tasks are scored basically on a four-point scale: definitely, probably, maybe, and unknown. These scoring categories indicate the degree to which a child showed a strength in a task relative to his or her peers in the classroom. (Scoring is discussed under "Condition 4: Clear Scoring Procedures.")

The scoring of the math worksheets and written stories is done by graduate assistants back at the University of Arizona. The math sheets are scored against a scoring sheet. Each correct answer is given a certain number of points according to the directions on the scoring sheet. For example, each correct arithmetic problem in the first set is given one point. The completion of each of the first two magic squares is given two points.

All four scoring categories for the traditional and alternative tasks are assigned relative to the class. To move from the points that are awarded on the math scoring sheets to the scoring categories the graduate assistants look for "natural breaks": The class papers are ordered from highest to lowest scores. Then, according to Maker, the graduate assistants look for "breaks," or score differences, between the papers of "four or five points that can distinguish the categories."

Maker stated that there usually are clear breaks in the set of papers that make it possible to distinguish the categories. "But if it's really hard to find one, then I suggested that people go back and look at the worksheets and see if they can see qualitative differences that might not have come through in the scoring. But usually that's not necessary."

The approach to scoring the writing task is, according to Maker, "holistic." There is no rubric or scoring sheet. Instead, a graduate assistant considers the "overall quality" of each piece and then divides a class' papers into the four categories. The assistant is then supposed to read through each pile to check that the papers within each pile are roughly consistent in overall quality. After that the rater puts the category on the back of the paper, mixes the piles back together, and a second rater repeats the same process. If there are disagreements between the two sets of ratings, it is almost always between two neighboring scoring categories. When that happens, the two raters discuss the papers they disagreed upon in order to reach consensus. On very rare occasions, when agreement between the two raters is not reached, a third person reads the paper and makes a decision.

THE ALTERNATIVE ASSESSMENTS

Along with the math and writing tasks, the DISCOVER process includes three tasks that are less traditional and school-like in appearance. These tasks do not rely on paper and pencil. The children tend to talk and even to collaborate during them. Their work is not collected at the end. Instead, members of the DISCOVER team observe and document the children's work on various instruments, and then draw on their observations to interpret and evaluate the students' work in later "debriefing" sessions.

These alternative assessments include the Pablo® construction activity, a tangram activity, and a storytelling activity, each of which has several components. The three are administered within a single day during a 2.5 to 3 hour period in the morning. There are brief breaks separating each activity during which the children are encouraged to get up and move around.

During these three activities, four to six children are supposed to be seated at one table with a single observer from the DISCOVER team. If there are no tables, desks are brought together and covered with butcher paper to prevent materials from falling between the desks.

The children are given nametags, which enable the observer to identify each child and record his or her work on two sets of instruments: The "Observer Notes" and the "Personal Interaction" sheets. There is one set of Observer Notes and Personal Interaction Sheets for each of the three activities. These are stapled into a six-sheet set for each observer. Each sheet is printed landscape fashion on 8.5 x 14 inch paper. The paper is organized into grids. Down the left side, the observer fills in the names of the students at the table. Across the top, columns are labelled with tasks or "comments" in which the observer can record the students' performances. Above the columns is a key of abbreviations which the observer calls on to document product and process characteristics. The column farthest to the right contains a checklist of five to ten characteristics, depending on the activity. (These characteristics are described below under "Observers' Role" and in "Condition 4: Clear Scoring Procedures.")[4]

In between tasks, while the students are taking their breaks, the observer moves to another table, organizes materials for the next activity, and gets ready to record the work of a different group of students. In general, each student is therefore observed by three different adults in the course of a morning.

At the beginning of the session, the classroom teacher is supposed to tell the class that there are guests visiting who want to do some activities with the class. Each of the DISCOVER team members then briefly introduces herself to the group.

When I observed, the introductions were warm and informal: Rogers, the team leader, told the youngsters at CES "...the reason we came to visit with you today, is to -- we've brought some activities that we've made up, and we want to see how you solve the problems we've designed. We're here to have some fun with you, and to watch you solve some problems." She explained that each table of students will stay in the same place, but the team members rotate tables after each activity. Each of the team members then briefly introduced herself. There was a conscious effort to establish rapport. For example, observers told students, "I'm glad to be here" or mentioned that they (the observers) are teachers who have gone back to learn more in school.

Each time observers rotate to a new table to begin a different assessment task, they reintroduce themselves and chat with the students. As one DISCOVER team member remarked: "It's a tone-setter type situation." Another said: "... getting started includes doing some of that interpersonal stuff, so the kids are comfortable with a stranger. You know, that, I think, is really important." After this, directions for the next task are given usually by the classroom teacher, and the children begin to work.

Pablo®

The first of the three alternative tasks is Pablo®. The task is named for the construction set that uses thick cardboard pieces cut into a wide range of geometric and free-form shapes, including circles, quarter- and half-circles, trapezoids, squares, triangles, and half-donuts, lollipop-like pieces, wavy lines, and teardrops. The pieces range in size from under an inch to more than six inches in length or diameter. The pieces use many colors: black, white, blue, turquoise, grey, red, yellow, orange. Some are solid colors. Others have designs comprised of contrasting colors, including checkerboard, concentric circles, stripes, triangles, and diamonds. Each Pablo® set for one table of students consists of 120 pieces and about a 75 "connectors." Connectors are black plastic pieces, about an inch in length. Each end of the connector has four spliced legs into which Pablo® pieces can be fitted. With some ingenuity, a single connector can be made to hold many different pieces, fit in at various angles to each other. (See Appendix D.)

After the team introduces itself, the Pablo® pieces without connectors are placed on the table. Then children are given six Pablo® tasks. The first is free-play. The second through sixth are said to move from closed to open-ended problems. After each of the six activities, the children are instructed to return all the pieces they were using to the center pile.

Task 1: The classroom teacher is supposed to read the following directions to the children: "You may take just a few minutes to make something with the pieces in front of you." This free play activity is intended to acquaint children with the materials. It is supposed to last about five minutes (Maker, Rogers, & Nielson, 1995, p. 4).

Task 2: The children are told, "The adult at your table [the observer] is holding a picture of a design. Make that design with the Pablo® pieces" (Maker, Rogers, & Nielson, 1995, p. 4). The design used was made out of construction paper and attached to a clipboard. It consists of a large square, inside of which is a circle, inside of which is a smaller square tilted at 45 degrees. The shapes are of contrasting colors. This task lasts about two minutes.

Task 3: The adult at the table picks up three Pablo® shapes: a parallelogram, a trapezoid, and an elongated hexagon. The children are next instructed, "The adult at your table is holding 3 shapes. Use two or more Pablo® pieces to make one of the shapes." (Maker, Rogers, & Nielson, p. 4). Approximately three minutes is allotted.

Task 4: The observer holds up pictures of a flower with spiky petals and a barrel cactus flower, that I was told would be familiar to the youngsters taking the test. The children are instructed: "The adult at your table is holding pictures of some flowers. Which pieces could you use to make flowers? Make your flowers on the table in front of you" (Maker, Rogers, & Nielson, p. 4). This task lasts approximately six minutes.

Task 5: The children each receive about a dozen connectors. Then they are asked to "Make something that moves with as many pieces as you need. Make anything that moves. You can tell us about it if you want to" (Maker, Rogers, & Nielson, p. 4). Approximately ten minutes is given for this.

Task 6: Students are encouraged to: "Make anything you would like to make using as many pieces as you want to use" (Maker, Rogers, & Nielson, p. 4). Approximately 10 minutes is also allotted for this activity.

Students' Experience of the Pablo® Activity

Throughout the Pablo® task, nearly all the children I observed were very engaged in their efforts and nearly all of them clearly enjoyed the activities. Children described the Pablo® tasks as "fun" even if the pieces are "a little hard to put together..." During the tasks, they talked, asked each other for pieces, and they occasionally volunteered pieces for others to use. A few times children collaborated to make large constructions, such as human figure of more than 20 pieces. A number of youngsters were very pleased with their constructions, especially in the free play and the last two tasks. They asked to have their work photographed. Many times, the children did not want to give back their pieces when a particular activity had ended.

Children's Pablo® constructions varied enormously. For example, the flowers ranged from a simple two-dimensional effort of two pieces, to another with 11 pieces, stacked into three dimensions with great attention to design and detail. During the latter two tasks, the constructions varied even more. There were small compact constructions of a few pieces representing motorcycles, spiders, mice, and other creatures. There were numerous large people, monsters, trains, and other vehicles. There were ensembles of constructions, including a man riding a bird, a mother holding a lollipop-eating child's hand, a man being shot in front of a target, a bear taking a cat for a walk. A few were conceptual or imaginary objects: a moving target that made its own bullets, and a "bumping thing" of pieces that bumped against each other. Some used 30 pieces, some used three, some were two-dimensional, others 3-D. Some clearly attended to detail: For example, for the mother and child with lollipop, both figures' heads were yellow, and their clothing was red. The mother's body was formed from two triangular pieces, outlining a women's narrow-waisted body. The girl's was a single semicircle. In others such attention was unclear.

The Observers' Role in the Pablo® Activity

Observers are kept quite busy during the Pablo® task. Each observer has to sketch on the Observer Notes all the objects each child constructs during the six activities. In addition, the observer looks for and records any of 19 product or process characteristics listed among the abbreviations or on the checklist section of the Observer Notes. For example, does the child attend to the design of the pieces, make use of negative space, work steadily? The observer is also supposed to write down comments that the child says about the work, among these what the object is or does. Alongside this, she keeps track of the students' interactions with each other and herself on the Personal Interaction Sheets. (See Appendix E for illustrations of observer documentation during the Pablo® activity.) The observer also takes photographs of the children and their work throughout the Pablo® activity, but especially during the last two tasks. These are used to document students' efforts. They are catalogued along with other records about the students' work and are used for research and training purposes. (See Condition 3: Evaluators are trained to carry out the work.)

Along with recording students' work in various ways, Maker reported that observers

> make sure that we're motivating the children and not saying something or doing something that would kill the motivation they have to do well on it. You know, if I have an expression on my face that says 'really, what Lee Anderson over there has done is absolutely incredible,' do I have that same look on my face for what Marcella did? And even though I might be able to control my verbal outbursts, did I raise my eyebrows, or did I nonverbally communicate something to them that might dampen their enthusiasm for it?

Thus, in addition to documenting students' products and processes in this and all the tasks, the observers have to maintain a warm, supportive, and equally encouraging relationship with the youngsters.

Tangrams

After the children have completed the Pablo® tasks and take a break, another observer begins to work with them on the tangram tasks. The observer gives each child a plastic ziplocked bag of plastic tangrams and asks the children to count the pieces to make sure there are 21. Each bag contains six large triangles, three medium-sized triangles, six small triangles, three parallelograms, and three squares. Children seated next to each other receive different colored sets to minimize the chances that a child will use a neighbor's pieces. The classroom teacher is supposed to provide the directions (Maker, Rogers, & Nielson, 1995) and give a brief demonstration of how different tangram pieces can be combined to make different shapes.

The teacher is to say "You each have a bag of colored shapes on the table in front of you. These shapes are called Tangrams. I would like you to take the Tangrams out of the bag. The Tangrams can be used to make many different shapes." Using tangrams of contrasting colors, the teacher demonstrates how a square can be made out of two triangles, a larger triangle can be made from two smaller ones, and a parallelogram can be made by attaching the shorter sides of two triangles. The teacher then demonstrates each of the following comments: "You also can trade or substitute pieces. For example, you can use a medium triangle to make part of the large triangle." "You can use a parallelogram to make part of the large triangle." "Finally, you can use a square to make part of the large triangle" (Maker, Rogers, & Nielson, 1995, p. 5). When I observed these sessions, the observers also demonstrated how to make shapes and substitute pieces for the children at each table.

After these instructions, the children are given two tasks:

Task 1: "Now make a triangle with as many pieces as you can." They are given about ten minutes to do this.

Task 2: The observers give each child a booklet of six green manilla pages. They are shown the booklet and told:

> Each page has shapes that can be made with the Tangrams. Be sure to make all the shapes on each page. When you are finished with each page, tell the adult at your table. She/he must check your work before you go on. Each page gets a little harder. Please continue working until you have finished as many pages as you can (Maker, Rogers, & Nielson, 1995, p. 6).

This activity lasts approximately 30 minutes. If a child finishes all six pages before time runs out (an unusual event), the observer gives out an additional, "challenge page."

Students' Experience of the Tangram Activity

In general, nearly all the students were engaged in working with tangrams. However, the atmosphere during this task was different from that in Pablo®. Rather than exuberance, children largely worked in a more focused way. As Aleene Nielson described the activity, "It seems to call for more concentration." Youngsters concentrated on seeking out the right combination of pieces to fill in the shapes on the harder pages. There was talking, and even some collaboration. Several times, children offered their classmates advice or help on which pieces to use where. The observers have a list of clues they can also give children, if the students ask for help. From time to time, children became frustrated as the tasks became harder. Some said that the page was "hard." A few said, "I can't do this." In general, the children persist with the task, even if they get stuck for quite a while on a single page. They do so partly because the task, though hard, is engaging, and also perhaps because the observers are quietly, but regularly encouraging them. (See below: Condition 2: Children are encouraged to do their best work.)

Observers' Role in the Tangram Activity

During the tangram task, the observer has several responsibilities. First, she acts as timekeeper: She records on the Observer Notes at what time each student ends each task and the order in which each child finishes each page relative to the others at the table. In addition, she notes on the Personal Interaction Sheets who helped whom, who looked to see how others solved problems, whether children asked each other for help, and any other kinds of interactions that may have occurred. The observer also records how children are solving the tangram puzzles: do they pick up pieces and set them down without rotating them? Do they lay the pieces over a form on the booklet and rotate them until the pieces fit? Do they try any pieces available, or do they have some sort of systematic search pattern?

Along with recording students' product and process characteristics, the observer must help the students deal with the challenges the activity presents. She does this in part through encouragement, saying, for example: "I know you can do this," or "that's right." The observer also sustains the students by giving two kinds of prompts. If a child works for about five minutes on a page, she can give the three "free prompts:" "you have enough pieces," "you can use more than one piece to make that shape," or "see which piece you can trade." These do not need to be noted on the Observer Notes for the tangram task. If the free prompts are not sufficient, the observers tell the children that they have clues to give, if the children want them. These clues range from "take the pieces off" to "use this piece here." When the observer uses these clues, she is supposed to record them in the Observer Notes.

Observers reported how they manage the challenges of this task:

> You have to be tremendously alert, tremendously keyed into the children. The moment I'm seeing any kind of indication that a child is feeling anxious, is starting to lose it, if you would, I go over there. I encourage. I use the three or more-prompts that we can give them....

Another observer voiced similar sentiments:

> The children express frustration often with the tangram task, where they're -- you know, as they get progressively more difficult. Some children will stop working and may use the tangram pieces to make a design [instead of completing pages in the booklet]. And I usually try and encourage children, saying -- acknowledge that sometimes things are difficult, and this is really challenging.... I use the clues that are given to help them. And if I see a child that's just super, super stuck with the tangrams, I think the last clue on the directions says something like 'use this piece here.' I do that, and then I note if I've given a child a lot of help.

Clearly, the children and the observers work hard during this task. It requires a great deal of concentration from everyone involved.

<u>Storytelling</u>

After tangrams there is another break, and then storytelling begins. For this task, children get a small ziplocked plastic bag containing seven plastic toys, ranging from about an inch-and-a-half to four inches in length. Each bag contains two different people (the possibilities include men, women, boys, and girls in different hues), two animals (e.g., horse, dog, vulture, cow, elephant), some sort of vehicle (e.g., motorcycle, car, truck, school bus), and two "things" (e.g., a telephone, a piece of furniture, a fence, a suitcase). When I observed that task, paper and pencils were on the table, in case a child wanted to write the story. Later interviews revealed that this option was taken away in order to focus the task on oral rather than written language. At each of the tables, the observers also have a tape recorder to record students' stories.

Storytelling includes three activities. The children are told "In all these activities, you may use any language you would like to use" (Maker, Rogers, & Nielson, p. 7). A bilingual observer or aide can work with the youngsters in the language the youngsters are most comfortable using.

<u>Task 1</u>: The children are told "Choose one of your toys and think of all the things you can say about it. Write these things on the paper I have given you, tell them to the adult at your table, or tape record them" (Maker, Rogers, & Nielson, 1995, p. 7).

<u>Task 2</u>: The children are instructed: "Now, choose 2 other toys and think of all the things you can say that tell about both of them. Write these things on the paper I have given you, tell them to the adult at your table, or tape record them" (Maker, Rogers, & Nielson, p. 7).

Task 3: The children are told to "Be thinking of a story to tell about some or all of the toys you have. You can tell any story you want to. Think carefully about your story. It cannot be longer than 10 minutes" (Maker, Rogers, & Nielson, 1995, p.7). These stories are told one-on-one to the observer, who also tape records each story.

Students' Experience of the Storytelling Activity

During the first two storytelling activities, most children seemed only mildly engaged. While some do give more elaborate information, the descriptions children provide for one object are usually brief. Of a toy woman, one girl said, "The lady likes to drive around in the car. She keeps her eyes on the road." A boy with a toy car said, "You can use it to drive to a telephone," A girl reported her toy car, "Goes fast. Has a good color. It has wheels. That's all."

The second task, in which the youngsters select two toys and are instructed to say "all the things you can say that tell about both of them," also led to short answers. These often indicated children's difficulty in interpreting the task as a request to provide features common to both objects. For example, in speaking of a car and a girl, one child said "They go to the store. They go to school together. They go to the mall." Of a monkey and a parrot, one child offered: "Fly, wings, long beak, climb around tree. Like to climb everything."

The story itself generated an enormous variety of responses, from virtually none to extensive, well-structured creations rich with detail. For example, at the table I observed in CES, all but one of the children was so involved in playing with the toys that they didn't want to tell a story. They saw the task as interrupting an otherwise good time. At the other end of the continuum is this example:

> It was in this one month, this man was driving in his car and he was going to a shoe shop for a new pair of shoes. His other ones were torn up. So he drives away over, down, and over the highway [gestures indicating car's movement]. Then he decided to take a shortcut through some woods, through the dirt road. Then after that, he kept on driving and driving. Then suddenly he sees this hawk in the sky. It's flying straight up, very smooth and fast. He kept watching and watching it. Going fast, like 30 miles an hour. All of a sudden a sheep was walking along the highway. He kept on driving, and finally he hit something. He hit the sheep. He felt something hit. So he rrmmrrmmrr [car sound] stops. Gets out of his car. He looks. He sees the sheep. He actually freaks out and says, 'Oh, shoot!' And then, after that, he decides to get the sheep, put it in his trunk and drive away to bury it somewhere. rrrmmrrrmmrr. Then he goes around the hill where nobody would find him. He had a shovel in the back of his car. So he decides to bury it. He started digging and digging, for up to six feet. After that, he just got the sheep, dragged its feet, and just threw it in. Then after that, he looked. Then after that, he just feels sorry for a while. Then he gets his shovel and just buries it back up. By the time he was buried, he makes himself a cross. He got two sticks, one was short, one was long. He gets some string, too. He ties it around so tight. Then, after that he sticks it in the ground. He hits it down. Hits it down with the top of the [unclear. Navajo word?] Like a hammer with the shovel. Then after that, and a while, he just stands there and took off his hat, and lay his head down, and said: 'I'm sorry about the accident.' Then after that he gets into his car and drives away. And takes off. Then a few years later, he's driving his car again. He goes to that place where he buried the sheep. Then, he went over there to see the sheep. The cross was still there. It was old. But it still stood up. He bent over there to pay his respect [sic] to the sheep. After that he said a few words: 'I'm sorry of what happened a year ago.'[sic] After that he just looked over and he walks away. At one moment in time, one last time, he really goes and looks again. After that he just get in his car [sic], turns on the engine, puts it in reverse, goes backwards, puts it in overdrive, and just drives away. Ssssshhhhh [car sound]. The end.

During the storytelling task, children who are not involved with the observer in telling their stories are, according to the directions, "encouraged to play with their toys on the floor" (Maker, Rogers, & Nielson, 1995, p. 7). There, or at the table, they often zoom cars around and crash cars into animals (likely a frequent occurrence in an area of unfenced grazing, and one which materialized often in the stories). In general, the youngsters create a noisy, happy, and even somewhat chaotic scene.

Observers' Role in the Storytelling Activity

Observers work one-on-one with each child for each of the three storytelling tasks The observer usually tape records the story or the children can opt to talk into the taperecorder. As with tangrams, the observers seek to encourage children to produce work, even when it is a challenge for them: Lee Nelson, an observer, described her approach to this:

> Storytelling may not be their thing at all. They may never have done any storytelling. And it's -- you know they're just not interested... And I just, I try. I don't force kids into that. I encourage them, and offer them possibilities. With regard to the storytelling, I'll say, 'if you'd like to write it first, you can do that or take the tape recorder over to the corner and do it. Or tell me the story, and I'll write it down.

The observers are again also keeping track of a variety of behaviors that the youngster is exhibiting. Aleene Nielson said:

> You often see younger children particularly putting their pieces together and creating, oh, stories with their actions. And so you can see some movement. You can see some leadership developing, as in who has the idea for the story and who's directing the story. You can see some of the kinds of interpersonal intelligence coming out when they're not doing something that's not directly involved with the observer. But you can sort of see it out of the corner of your eye.

For storytelling, then, as with the other two activities, the observer has a multifaceted task: recording actual work, recording children's behaviors, and interacting in supportive ways with the children.

Evaluation/Scoring of the Alternative Assessment Activities

The evaluation of the Pablo®, tangram, and storytelling activities is carried out by the team of observers on the same day that the tasks are administered. After the storytelling task, the observers move to a quiet area in the school. Each of them then extends information that they did not have time to record on the Observer Notes and Personal Interaction Sheets during the actual assessment.

A third instrument, the "Problem-Solving Behaviors" checklist ("the checklist") is sometimes also filled in at this time. According to Nielson, the checklist is supposed to be filled out for each student after the observers' discussion of all the students' performances. The checklist consists of nine stapled 8.5 x ll inch pages for each individual student. The checklist pages are organized around different intelligences: linguistic, spatial, logical-mathematical, interpersonal, intrapersonal, bodily-kinesthetic, plus one cross-cutting category, called "general." The latter includes behaviors which the designers do not link to particular intelligences, among these: "persists on tasks that are difficult for him/her," "attends to own work," and "organizes materials."

The checklist looks at products and behaviors within an intelligence as this intelligence is employed across tasks. Thus, for example, listed down the pages devoted to linguistic intelligence are characteristics such as "tells stories easily and fluently," "uses more than one language," and "chooses colorful or unusual adjectives and adverbs." Across the columns at the top of the page are the different DISCOVER tasks in which these behaviors may have been manifested. The DISCOVER team culled this list of behaviors by watching youngsters as they solved problems, writing down what they saw, and then discussing those behaviors "that indicated superior problem solving in that activity."

The checklist for each child circulates among the observers. Thus, the observer who worked with a child during tangrams checks behaviors that the child manifested in each of the intelligences during the tangram activity. Likewise for Pablo®. Graduate assistants who score the storywriting and mathsheets back at the University of Arizona also complete the checklist.

As they work through the Observer Notes and Personal Interaction Sheets, the observers may make some tentative determinations about youngsters who showed strengths in the tasks they observed. When observers have completed extending these two instruments, they begin the "debriefing session," during which they discuss and compare students' performances and determine the youngsters' scores.

The discussions proceed serially through each of the three activities. Usually, scores for all the students on Pablo® are decided before beginning the scoring of tangrams. Storytelling is evaluated last.

At the beginning of the discussions for each activity "we're sort of establishing the criteria for the classroom," Rogers said. That is, the observer team tries to figure out, for the activity under discussion, where the four scoring categories map onto the performances of the students they just observed.

They begin by trying to decide where the "definitely" is. Typically the discussion for each of the three activities begins with a statement like "What is your definitely?" or "Did anyone have a student that indicated an unusual strength in this particular area?" Then the work of that student is discussed and compared with other potential "definitelies." Students whose work is not quite as strong are designated probably; work that is less strong than probably is designated maybe. The "unknowns" are supposed to designate students who "just didn't do anything" or whose performance was limited to an extent that his or her strength in the task is unknown.

Youngsters who get a definitely in two or more of the five tasks are identified as gifted. (The ratings for the two more traditional tasks are added to the class list and children's checklists by the graduate assistants back in Tucson).

The students' efforts are discussed along a number of dimensions, including but not limited to ones designated on the Observer Notes, Personal Interaction Sheets, and the checklist. (Further details on the evaluation of the work appear in the section below in Condition 4: Clear Scoring Procedures). One child's story and its evaluation may occupy 15 minutes or more. Given this, the debriefing sessions extends over several hours. The sessions I observed lasted between 3.5 and 5.5 hours, for classes ranging in size from 18 to 27 children.

ANALYSIS OF WHETHER INCREASED IDENTIFICATION OF UNDERREPRESENTED YOUNGSTERS CAN REASONABLY BE ASSOCIATED WITH THE DISCOVER ASSESSMENT AND WITH MI

In Chapter 1, I described five general conditions that are needed to associate increased identification of underserved students with the assessment efforts I am investigating. These conditions need to be met to make inferences about any student from any assessment. I also described three conditions that should be present in order to associate the assessment with MI. In the following section, I analyze whether each of these conditions is met. When the condition is not met, I offer suggestions as to how the assessment might yet be strengthened.

GENERAL CONDITIONS

<u>Condition 1: Children Understand the Tasks</u>

In general, children taking the DISCOVER assessments do understand what they are being asked to do. Efforts on a number of fronts help to ensure this:

First, children are supposed to be instructed, and work with, an observer who speaks the same language as the children (Maker, Rogers, & Nielson, 1995). Rogers reported that allowing children to receive directions and work in their native language is "such an innate part of us that we frequently forget to tell people" about it when describing the assessment process. The execution of this DISCOVER principle, like most principles, sometimes falls a bit short in practice. Until recently, teachers were supposed to give the DISCOVER team advance notice of the language needs of the students. However, teachers either did not always know or did not always report to the DISCOVER team their students' language preference. To correct for this, the children are now told in the beginning of the assessment that they can use whatever language they are most comfortable using. To support work in the children's preferred language, the assessment team includes people who are fluent in Spanish, and they have drafted Navajo graduate assistants. The team has also trained Native American teachers at the schools to participate in the assessment process. I observed a Native American teacher working with one group of students at Chinle Boarding School. However, at another time, that teacher was not available. In this case, the observer team was assisted by a classroom aide who was a native speaker, but not trained to work with the DISCOVER team.

A second route to ensuring that children understand the tasks, according to Maker, is to keep the directions simple and concise. As the above descriptions reveal, directions for the tasks are conveyed in a few short sentences. As those descriptions also illustrate, the directions are often supplemented with examples and demonstrations.

While the directions are generally clear, there are two areas of ambiguity. In the fifth Pablo® activity, when children are given connectors and asked to make something that moves, some youngsters depicted things that could move, e.g., vehicles or animals. Others made objects that actually did move, e.g., a figure with connectors for hips that did splits when it was pressed against the table. Observers tended to be quite impressed by contructions that actually moved. If a figure that moves is the desired outcome, this should be made more clear to the children, perhaps with a working demonstration. If it does not matter whether children depict actually movable objects or not (Nielson, personal communication, February 18, 1997), then the observers should not score such constructions higher than immobile constructions (See Condition 4: Clear Scoring Procedures.).

In addition, the storytelling activity contained one set of directions that children consistently did not understand. The second storytelling task asks children to "... choose 2 other toys and think of all the things you can say that tell about both of them" (Maker, Rogers, & Nielson, 1995, p. 7). As noted earlier, children's responses to this revealed that they often did not understand what was being asked of them. In the debriefing session for Chinle Elementary School, Rogers noted that one girl clarified the directions for her: "'Oh, you mean what they have in common.'" Rogers agreed, "Yes, that's what I mean." However, the directions aren't given in this way because "not all children know what I mean when I say 'what they have in common.'" Therefore, Rogers gives the children a prompt, "they [the toys] both ..." However, as her teammates noted, this prompt is not in the directions, and the other team members did not use it.

As Rogers' exchange with the student reveals, the observers' interactions with the children is a third way of fostering children's understanding of the tasks. Another observer's comments also reveal how such interactions support students' understanding of the tasks:

> And when I give directions, I look at children to check for understanding. ... [They] ask questions, you know, 'Can I--?' 'What can I do--?' 'What am I supposed to do?' 'Could I--?' 'Can I use all the pieces?' You know, they usually ask questions that help me to clarify what it is that they can be doing.

Of course, occasionally, individual children did not seem to grasp particular tasks. For example, in the first tangram activity, some children made a square instead of a triangle. Even after prompting one child with a comment about how nice the square was and asking the child to go on and make a triangle, the child did not proceed. However, in general, nearly all the children understand nearly all the tasks. In the Pablo® and tangram tasks, the vast majority of the youngsters worked steadily and produced products that mapped onto the directions. For example, the overwhelming majority made flowers and understood to use the connectors to attach Pablo® pieces. For another example, nearly all of the children attempting to place the tangram pieces correctly in the outlined shapes.

Given the helpful interactions, the clarity of nearly all the directions, and the opportunities for children to work in their native language, it is reasonable to say this first condition is met: Children do understand the tasks.

Condition 2: Children are Encouraged to Do Their Best Work

In general, DISCOVER assessments meet the condition of encouraging children to do their best work. This is accomplished in a variety of ways:

First, at least in the three alternative assessments, children are provided with materials that are interesting to them. As observer Lee Nelson put it, "One of the things about the tasks is that they just seem to really engage kids right away." Aleene Nielson said:

> I think one of the things they see is that they're being asked to do some things that are fun ... the materials are brightly colored, they're engaging.... some materials they've probably not seen before, and they look like toys. I've even had adults say how much fun it was to do the Pablo® tasks.

A second way students are supported to do their best work is by minimizing the language demands within the assessment. As noted above, directions are simple, and students are encouraged to work in the language that is most familiar to them. In addition, students have a diversity of materials with which to demonstrate their thinking and problem solving: Their strengths need not be demonstrated primarily via language. According to Rogers, "We've tried very hard to not make language be a barrier for any child's abilities -- in other words, not let language get in the way of allowing children to show us what they're capable of doing." (This quality is discussed further in the discussion under Condition 6: Intelligence-Fair).

Children are supported psychologically and emotionally to do their best work. As noted above, the observers foster rapport with the children. This happens both in the morning when all the team members introduce themselves to the whole class and each time the observers rotate to a new table. There, as one observer described it, she spends "some time visiting individually with the kids -- you know, just saying 'hello' and 'how are you?' And asking a few questions, answering questions about what's going to be happening, to set the scene for students..."

Relatedly, children are not told that they are about to take a test. Rather they are told that there are some fun activities that the visitors want to do with them. In addition, the children remain in their familiar classroom environment, with all their classmates. "We don't pull kids out into a site that is unfamiliar to them," said Nielson.

In this familiar setting youngsters are given encouragement to facilitate their best efforts. As reported earlier, students are attended with equal enthusiasm and encouragement. They are told things like, "I know you can do it." They can seek clarifications from the observers, and in the tangram task, the observers can provide them with clues to help them.

A possible exception is that the DISCOVER alternative assessments may be less successful in encouraging the best performances of students who are shy. This exception was noted by both teachers and designers.

The two, more traditional tasks are somewhat less strong in the area of encouraging children to do their best work. For example, the teachers are not supposed to help the children with story starters. In addition, the nature of the tasks encourages children to follow the school script: work by yourself, without other resources. To support students' best work in storywriting, students could be encouraged to try out stories with their friends before writing them down. They could be told in advance that they're going to be asked to write a story the following day so that they might think about, and talk to others about, the task. Although the two traditional tasks could be revised somewhat, overall the spirit and implementation of DISCOVER does fulfill this second condition.

Condition 3: Observers/Evaluators are Trained to Carry out the Work

While evidence indicates that DISCOVER III meets the first two conditions, the same cannot be said with regard to the training of the observer team members. Part of the problem in meeting this condition is the fact that the team members have a complex role. Their charge is to observe, equally attend, and record in various media the products, problem-solving processes, and interactions of several, often very busy youngsters. (See Condition 4: Clear Scoring Procedures). The observers' role also includes evaluating youngsters' performances in the alternative tasks. The work presents many challenges, even to those most highly experienced with the process.

Rogers commented that:

> Probably the key thing, the key components in any observer's mind, if he carries any salt whatsoever, would be, you know, 'I have to make sure I'm getting this down [recording the students' effort].' This is something key: 'Look at the interesting method that child is using. That's important. I have to put that down.' In my mind, those are the kinds of things I'm thinking. And then ... there's also the rooting kids on, you know. Act positive: I know that some kids are getting through those pages [of tangrams] simply on the positive reinforcement our observers are giving them.... And the challenge of making sure that you're adequately observing all students and not getting totally overwhelmed by some student who is doing such a super fantastic job that you neglect to see the wonderful things that the other students are doing.

Maker reported:

> One of my hardest things is when we're videotaping a group and I'm trying to 'woman' the video and observe my kids, and take all the pictures, and sort of be all those things at once, and the video camera stops blinking. So I have to figure out what's going on with it. That's like a major task. But if I'm not dealing with the video, one of my major challenges is just how do I possibly write down everything that I see and capture it? I mean, I

> know that I'm getting more through this kind of assessment than I would if I were just giving a test. But it's so rich with things that I could take down, 'am I missing important things?' is always a question going on in my head.

As these statements illustrate, DISCOVER observers' work goes far beyond the timekeeping and cheat-detection functions demanded of proctors during traditional tests. Clearly, the DISCOVER III observers require training and practice.

To DISCOVER's credit, Maker and her colleagues have evolved an intensive training effort which lasts a minimum of three days. This training is typically provided to people outside the nine DISCOVER sites who wish to learn about and apply DISCOVER in their own schools. A slightly modified version of the training has also been used with the graduate assistants who work with the DISCOVER observer team.

The training generally follows this format: On the first day, participants get a brief overview of the theoretical origins of DISCOVER and its approach to assessment. Then, they take each of the alternative assessment tasks, so that they appreciate the kinds of obstacles the youngsters encounter. They also practice administering the tasks. In addition, they look at and discuss slides and videotapes of student work. In these discussions, Maker said "we would talk about what are the characteristics of a particular child's products: You know is it three-dimensional? Is it complex?" After this "of course, we tell them what we think ... [based on] the characteristics of products that are included in the ... behavior checklist." Trainees also look at slides and/or video of the work of a table of children alongside the observer notesheets for that group. This gives them "an opportunity to see what an experienced observer would write down in response to what they saw."

During Day 2, trainees are paired in a practice session administering the assessment in real classrooms. Maker said the pairing enables trainees to "learn from each other" and makes the first assessment experience "not as overwhelming." Nielson reported that graduate assistants who work with the DISCOVER team are paired "with an experienced observer for the first time or two. Because that allows us to build on the inter-observer reliability and to increase their knowledge base..." Ordinary trainees, Maker reported, may not have an experienced observer in the room to consult. After this, the pairs of observers go through a debriefing discussion. They discuss what they saw and present it to the other trainees. They then complete the students' checklists, asking questions along the way to clarify items on the checklist.

The third day entails practicing how to score the math worksheets and the written stories. These have been administered and collected by the classroom teacher from the students that the trainees observed the previous day.

Although a training process has been devised, it is not always used to train the DISCOVER observers. Rogers reported:

> [with] the new hires, what I had them do in the very beginning was to first watch our introductory kind of video and parts of our assessment video. But then I actually had 'em take observer notesheets and the video that went with that, and actually sit down and watch the entire thing ... [I]nstead of a big training session for these very bright individuals that I've hired, that has worked very well. They've been able to go from that, and just sometimes [go] into doing an observation that has been a very good observation. I wouldn't do that with any Tom, Dick, or Harry, but --.

Thus, some DISCOVER trainees have an extended training, and other, very capable newcomers get an abbreviated version. One of the latter individuals said, for training, she "watched a few of those videotapes, and then I, I was just launched into doing assessments. I just started doing them. Didn't observe anyone doing an assessment, nothing like that. I just started doing them." This observer commented that observing is "just tremendously complex" and that during her "first couple" observations she felt "pretty inept."

> [M]y initial experiences were just really overwhelming, because there's so much to do in terms of management: getting materials out, and then you have -- you're observing. If you're working a videocamera, that has to work. And you're supposed to be taking photographs, keeping track of time. So, there's a lot of --. It greatly resembles learning how to drive a car. You know, so a lot of things are much more automatic now.

Another observer recalled that she had been drafted into observing, even before being hired or trained. This occurred when she was invited to see a DISCOVER observation, while considering whether to work with the DISCOVER team.

> MK: Can you describe how you were trained to carry out the assessments with the students using these tasks?
>
> Observer: First, I was given the written information. No. That's not true! I was thrown into it when I went to Nogales on that visiting trip.

Teachers at the sites may also not have the level of practice needed to conduct an adequate observation. At CBS, the teacher who worked with the Navajo-speaking youngsters had been trained, but she had not participated in many observations. Rogers noted: "Many of the [CBS] staff have been trained to do the process. I feel that after watching [teacher's name] today, that they might need some refreshers, because a few things might have changed. They need maybe a guided experience." Susan Bartley, director of the enrichment program for Chinle Unified School District, sometimes also participates in the observations. She expressed sentiments in line with the idea that educators at the schools who serve as observers may need more practice than they receive:

> I think if I did it [observing] more, I would know exactly what I wanted to know [about the students' performance]. But it's just all so interesting, that I want to know it all. And it's really hard when these kids won't stand still long enough. You know, they get excited. You get excited. Before you know it, you forget to -- you can't remember what some of the kids did.

One of the reasons that inexperienced visitors or less practiced school personnel may be called upon to observe is that everyday events intercede while the DISCOVER team is visiting. Sometimes, a class size turns out to be larger than reported and extra observers are necessary. When the team is observing, they often travel long distances (it is about a 7-hour drive from Tucson to Chinle) and put in days that last 10-12 hours. Given this, observers occasionally become ill. Another reason that unseasoned observers participate is that the designers have great faith in the people they hire. As Rogers said, they don't hire any "Tom, Dick, or Harry." Among the less experienced observers I accompanied to the reservation were two seasoned elementary school teachers, both about 50 years old, who had left the classroom to work and to study.

Finally, it appears that novice observers participate because such participation is seen as part of the training process. As one observer put it, "in terms of learning how to participate in the debriefing, it's by participating."

No assessment team is every made up totally of novices. When I was there, I was used as an observer. I had been through a three day training, and two observations of the process. Two other people were relatively new as well, having participated in seven and three assessments. Two others had considerably more experience. Several times, I was rescued from potential mistakes by Rogers, who was able to monitor her own group and make sure that I was not going astray. Some errors no one could save me from. For example, I made up some clues to give to children during the tangram task.

Yet, because the information gathered during the alternative assessments is central to students' identification, and because students' can take the DISCOVER assessment only once or twice a year, it is important for this information to be complete and accurate. Thus, it would be beneficial to ensure that the observers are highly skilled *before* they participate fully in this complex work. One way to accomplish this would be to have new observers serve as apprentices. They could assist with the photography or video work, thereby alleviating some of the burdens mentioned by the experienced assessors. They could also listen in on the debriefing sessions, gaining some skill in the evaluation process. These or other approaches are needed to strengthen DISCOVER's identification process. At the present time the evidence is not strong enough to say that observers are sufficiently trained to carry out their demanding role.

Condition 4: Clear Scoring Procedures

In the description of the DISCOVER process, I outlined the procedures used to evaluate student work: Each of the five tasks is usually scored on a four-point scale: Definitely, possibly, maybe, unknown. A child who receives a definitely in two or more tasks is identified. The mathsheets are evaluated by graduate assistants in Tucson who assign points according to a scoring guide. The written stories are scored independently by two graduate assistants who judge the work holistically.

For the alternative assessments, the scoring occurs during the debriefing. As noted above: "the usual first question is: 'Did anybody have a student that indicated an unusual strength in this particular area?'"

> And then somebody in the group will say, you know, 'I had so-and-so.' And I think what is established, although it's never really been stated as such, but what seems to me is we establish sort of a benchmark for each of the designations [four scoring categories]. There may be another student in another [observer's] group whose work really in comparison ends up making my ['definitely'] student a probably. Because as we discuss and give specific examples of what we saw and what our notes seem to indicate, there's often a child who's just far exceeded what other children have done.... Once the [definitely] has been established, then it seems to proceed from there.

Usually the observers then work down the scale from definitely to students whose strengths are "unknown." Sometimes, they may jump from the definitelies to the "unknowns," noting children who "just didn't do anything." For the other categories, "probably" and "maybe," "it's just a matter of discussion, discussion and adjusting." As Susan Bartley described it, the members of the assessment team:

> sit down and talk about the kids until we came up with a consensus of what we saw. And it was consensus. It wasn't one person saying, 'this is what I see and this is the way it is.' It was consensus on -- I still think that's pretty open to interpretation. Sometimes I'm kind of uncomfortable with it. Sometimes I see something that I think is really unique, and other people don't see that.... [But] We come [up] with a consensus. I mean, I usually -- I don't always change the way I think, but I can see their thinking.

<u>Establishing evaluation criteria for the three alternative activities</u>

As Bartley's comment suggests, while the structure for the debriefing sessions is clear -- discussion leading to consensus -- the criteria used within the debriefings are much less clear. As described below, the three alternative activities vary with regard to clarity of evaluation criteria. After considering each activity, I discuss issues affecting the scoring of all three.

<u>Tangrams</u>: The criteria for evaluating tangrams are "the easiest to establish," according to Rogers. Another observer team member supports this: "That's pretty easy

because, you know, if you have somebody that goes through all the tangram puzzles and is on the purple sheet [the additional challenge page], then 'HELLO?!'" A discussion with team members during a debriefing session again reveals that the criteria for judging tangrams are relatively clear:

> Observer 1: With tangrams, I think you can say when you get to [page] 6, that's a definitely.
>
> Observer 2: Uh-hm.

As these comments indicate, the number of tangram booklet pages that a child finishes is a strong criterion for evaluating this task. However, the number of finished pages was tempered by other considerations. Sometimes, the speed in working on the pages influenced scoring:

> Very generally speaking then, we're kind of going by completing the book as being definitelies. However, I'm giving [child's name] a definitely, because of the fact that she went through all the other pages very quickly, even though she didn't get page 6.

The amount of time a child actually devoted to the effort influenced the scoring as well:

> ... I feel very confident that R would [have finished] ... But a lot of that time she spent working with T to get [T's pages] 3 and 4 finished. ... [I]nterestingly enough, at the end, it wasn't hers that she was continuing to work on. It was with C. She was trying to help C after time was called ... rather than doing her own page."

The amount of help a child received from peers or the observer also influenced tangram scoring: "I don't feel comfortable with anything but a maybe, even though he was working on page 6. Because he did receive help on both page 3 and page 4." Another example shows how the observer takes into account the clues she gave "...we went through all the clues. And I finally went through Clue G with him. So, he's the one I'm thinking is a maybe. Because he needed that help."

When it was unclear which category to apply, the presence and size of the beginning triangle helped tipped decisions: "I know the other reason I didn't feel like I could give S a probably: He didn't do a triangle."

Occasionally, whether a child evinced problem-solving strategies in doing the work also influenced decisions. For example, a child was given an unknown, even though he had done some pages, because:

> ... He didn't seem to demonstrate any strategies, like trading, or moving or flipping [the pieces]. I watched him work with the pieces and having two [individual] triangles, [he] spent a very long time trying to make a large triangle [from the 2 smaller ones].

In sum, for the tangram task, the major criterion for scoring was clearly the number of pages completed. The amount of time actually spent doing the tangram activities, the amount of help received, production and size of the initial triangle, and evidence of strategies also influenced the decision.

While these criteria were the overwhelming ones used by the observers, some ten other process behaviors associated with tangrams are listed on the Observer Notes. These include, "organizes materials," "continuously working," "encourages others," "chooses shapes without turning." There are also ten, somewhat overlapping behaviors on the Personal Interaction Sheets, among these "organizes group activities," "encourages others to try," and "competes with others." There are 86 somewhat overlapping process or product characteristics that could possibly be checked for a child's performance on tangrams, including "directs the spatial component of a group effort," "others listen and respond to him/her," and "demonstrates high-level eye-hand coordination." These were rarely mentioned. In nearly all cases, discussions focussed on just the few criteria mentioned in order to achieve consensus about each child.

Pablo®: The key criterion for strong work in Pablo® is "complexity." Without complexity it is very hard for any child to be designated definitely. But what is complexity? To Rogers it was evident:

> In my mind, Pablo® is really pretty clear also. If you have a student that consistently makes complex, three-dimensional structures, whether it's symmetrical or asymmetrical, done some problem solving, I think that's a pretty strong [performance]. I know it when I see it. Kids who put things together in an intricate manner. To me that would be complex.... If its's just put together with four little [pieces], that's not complex. If a lot of pieces are stuck into one connector, [pieces with] various shapes and forms -- and some kids can do it -- that leads to complexity.... You know complexity when you see it.

Interviews and observations of debriefings confirmed that complexity was important. However, as another observer revealed, perceptions of complexity varied, based partly on observer experience.

> You know, how many Pablo® constructions have I looked at now? And so what seems to me to be complex in the first one that I did, probably is not going to be complex. I think if I went back to the first assessment that I did, and looked at the assessments that I'm going to be doing in the spring, I know that my experience makes what I do now as an observer significantly different.

Comments from the debriefing sessions reveal that the presence of three-dimensional work using the connectors contributed to the attribution of complexity. Three-dimensional products offered a dividing line between "definitely" and lower categories. For example: "And she made a pizza. And it was nothing. Nothing was there. It was all nice stuff, but nothing really spectacular. Not complex, nothing three dimensional. ... So that was my 'maybe.'" For another example:

> Observer 1: ... What was the rollerskating person?...
>
> Observer 2: It was very simple. It was the big, the big hexagon shape. The big triangle on the top of it for the head. Two rectangles for arms, and the lollipop pieces were the legs.
>
> Observer 1: Nothing kind of going out [in three dimensions] or anything?
>
> Observer 2: No, no, no, no. One very --
>
> Observer 1: Two-dimensional.
>
> Observer 2: Right. You could lay it on the table and it would be flat, even though the connectors were there.

Unless one or both of the last two Pablo® activities used a variety of pieces, and employed 3-D, the student's product was not seen as complex and was not accorded the "definitely" needed to support gifted identification. This point is highlighted in the presentation of the work of a child who did attend to symmetry, shape, and detail, though not 3-D:

> He picked and used almost throughout, the arrow pieces and the L-shaped or V[-shaped] -- whatever those are. And to begin with, he used them in just this negative space.... Then for his flower, he took the pizza pie pieces and ... he put all of those around and they were all going in the same direction. Then when he came to things that moved, he made five little one-piece things. He did not say what they were. But once again, he took them and put the head -- what would appear to be the head -- and put them in that order. And then he took the wavy piece and two of the Ls and two of the arrows on top of that. And he did call that a something, and I've got it: A yaahbichii [a traditional Navajo figure]. He called that a yaahbichii .... But everything he did had those arrow pieces in it. And everything he did had the four [v-shapes] -- he made his flower with those things like I said.... There was nothing three-dimensional that he created. I said he was a probably.

While complex and 3-D structures were necessary for a "definitely," these characteristics were not always sufficient, as the discussion below illustrates:

> Observer 1: And then it's holding a pizza, and the pizza had all the parts on it. And it was very complex. But it was very -- looked a little messy to me. She would've liked to have filled it out. But she spent a great deal of time on it. She paid real close attention to the design on the pieces and kept reworking them for just the right effect. Because she had several shapes that were appropriate, but she didn't care for the color.... She seemed to have a definite color scheme and shading in the face that she was going for. So, I thought that was enough to qualify her for a probably in this group.
>
> MK: Why isn't it a definitely?
>
> Observer 1: I didn't think that this was complex enough in view of what you all were saying. I thought -- the thing about this was, it was complex, but it started looking like a large pile. You know, they say good art -- sometimes artists don't know when to stop. You overpaint or oversculpt. And it looked like this to me. The face, I could hardly go through the layers. And it just didn't -- it didn't seem strong enough.... I mean there were strengths in there. That's why she's a probably. But I did not see -- I didn't see that something that lets me know that, that it was [definitely]. I thought the ceiling was very high in this class.

As this quote underscores, the scoring of all the work is relative to each child's classmates. (This issue is considered at the end of Condition 4.) Therefore, in classrooms like the one in CES, where many children did complex, 3-D constructions, other aspects of the work were highlighted to distinguish the quality of one child's work from another's. For example, the observer above felt the work was overdone, "messy." In another case that entailed scoring two girls' collaborative construction, the observer team considered whether the girls participated equally in the work, whether the observer should rely on her own "gut reaction" to the two girls' efforts, and whether to give one child in the pair a "definitely," based on "the benefit of the doubt" -- a principle widely applied in ambiguous cases. Messy, overdone, gut reaction, and benefit of the doubt are not listed among the somewhat overlapping 19 product and process characteristics on the Observer Notes for Pablo®, the ten on the Personal Interaction Sheets, or the 78 items in the Behaviors Checklist. Yet, they are drawn upon, while many process and product characteristics listed on the instruments were not used, among these: "competes with self," "demonstrates confidence in self," "construction demonstrates knowledge/ understanding of self," "invents and plays with words." In fact, it is not always clear to the observers whether, for example, they should draw on given criteria for Pablo®, like "organizes group activities" because this speaks to the child's interpersonal, rather than spatial abilities.

In sum, when complexity/3-D is a sufficent basis for scoring, the scoring of Pablo® is reasonably clear. When complexity/3-D is not sufficient, the criteria for scoring appear broad and open. Though the observers achieve consensus in scoring, it is not always clear on what basis.

Storytelling: The picture emerging across the first two tasks, is that though there are many possible evaluation criteria, the criteria actually drawn upon to delineate "definitely" from other scores were, for the most part, clear. In both schools, the observers relied heavily on the number of pages completed in tangrams, and the presence of complexity/3-D in Pablo® to make their decisions. In contrast, storytelling criteria were much more variable and harder to identify. Rogers felt that the criteria for scoring storytelling were also clear. Yet, her comments also reveal ambiguity:

> And story, really the criteria I think is [sic] pretty clear, too. I think when all of us read our stories out loud, you can see the difference between a definitely -- it hangs together, there are a lot of story elements. There's a sense of that [story] being something that contains more than something that's not really a story. Does that make sense?

Unlike the two other alternative tasks, data from the two different schools did not converge on two or three central scoring criteria for storytelling. This may be due to the fact that the children in the different schools produced markedly different "stories." In CES, many children produced sophisticated stories, such as the one highlighted earlier about the man who killed the sheep. These stories sparked comments by observers that went beyond whether the story "hangs together," to more subtle story elements: the use of dialogue, vocabulary, irony, humor, an apropos conclusion (not merely the presence of a conclusion), and whether the story was fully realized -- whether a child "could've done more" with it. In contrast, at CBS the observers focused mostly on the presence or absence of a story structure and whether the characters' actions could be traced to any motivation.

As with Pablo® and tangrams, there are extensive product and process characteristics for the observers to consider in evaluating storytelling. The Problem Solving Behaviors Checklist includes 81. Of these, some 20 are story-pertinent characteristics, e.g., the use of "complex sentences or syntax, "a sequence of events that is appropriate to the story," "chooses colorful or unusual adjectives and adverbs." Many of these are overlapping: (e.g., "stories have a recognizable plot," "stories have a recognizable beginning, middle, and end," and "stories have a sequence of events that is appropriate to the story"; or, "stories include complex and/or sophisticated words or concepts" and "stories include complex ideas (e.g., philosphical, moral, spiritual, political, cultural)." At the same time, there is little overlap of task-relevant characteristics across the Observer Notes and Behavior Checklist. This detailed and extensive approach is too much for observers to juggle. Ironically, it renders the bases for decisionmaking unclear and underdefined.

One observer expressed what emerged for her as the criteria for judging stories:

> It almost seems when I listen to people [i.e., the other observers] what jumps out is what people use as the basis for [judging] each individual piece. So, it's almost as though, if you have a really strong feeling about the way the language flows, then that's the reason. And then for someone else, the reason might be because there's a beginning, middle, and an end. Or, the piece that we heard yesterday, because we thought it was sensitive and came so much from within. And yet, structurally, it was lacking considerably. So, you know, I think there is an individual approach to each piece. And maybe that's the way it needs to be. But you really, you can't -- I wonder if you actually can't compare, for, you know [some characteristics]. Otherwise, I think you'd have to have some sort of performance checklist: O.K. It has a plot: check that. It has a voice: check that.

Though criteria for it are the least clear, storytelling is the only one of the three alternative tasks for which criteria can be gleaned from an existing, real-world genre. Real stories do need to hang together, as Rogers mentioned. But alongside a coherent plot, one could look for setting, characterization, use of detail, and other elements intrinsic to constructing a story.

Alongside unclear criteria, the task itself (storytelling versus story constructing) is unclear to the observers. For example, one child simply retold the story of Jack and the Beanstalk. Because neither the child nor the evaluators knew whether retelling a story fulfilled the task, some ten minutes was spent considering whether the criteria for judging this effort should include "the quality of the voice" and "the emotion" used, or whether this retelling was basically plagiarism.[5]

In sum, for storytelling, unlike the two other alternative tasks, no clear set of evaluation criteria emerges.

Issues Pertaining to Scoring Procedures Across the Alternative Activities

Several issues relating to clarity of scoring criteria apply across different alternative tasks. This section examines these issues.

"Many called; few chosen:" Across the alternative tasks, there is a vast number of process or product characteristics that observers can call upon from the three different instruments they use to evaluate students' efforts. In fact, the list is open-ended. New criteria are invited by blanks and boxes on the instruments labelled "other" or "comments." Thus, comments like "messy" can then be drawn upon and considered. The upside of this is that it allows the observer to note what is special or important about what she observes in a student's work. The downside, when considering the clarity of the scoring, is that virtually anything can influence observers' judgment.

In fact, however, few criteria are called upon in the first two tasks. One reason for this may be that, in a large, complex, and time-bound task like the DISCOVER debriefing, it is just too difficult to consider all the possibilities. Instead, evaluators "satisfice" (Simon, 1979): they accomplish their task by relying on a limited amount of information.

Given that Pablo® and tangram scoring is heavily determined by a few criteria, it is unclear why so much other information is listed and collected. This practice adds considerably to the time evaluators spend scoring the work. Further, not all the characteristics that are listed are fully understood: There were discussions over the meaning of "alliteration," and "negative space," even among experienced observers. There were discussions about whether one instance of a behavior should lead to a check, or whether several instances were necessary. Because many of the characteristics are little used, and some are unclear, paring down and clarifying the items would enhance the scoring procedure.

"Escaping the bear:" Another challenge observers face in scoring all the alternative tasks is maintaining the classroom-based standard or reference group. Maker argues that comparing children across classrooms is problematic:

> [W]e can take what happened in that classroom that year and say, that for that group of kids, they probably had a fairly similar experience. Once they get out of that classroom, we can't say that. And so, to me it's much more valid to evaluate them within that classroom context than it is to try and compare the kids who are in that classroom to the kids who are in a different classroom.

Another member of the observer team, Claudia Clark MacArthur, said that when she had asked why the scoring parameters were set at the classroom she "was comfortable when the answer was given to me: 'because that is their environment and the expectations are that they have received the same opportunities.'"

As fundamental as this principle is to DISCOVER, observers cannot simply obey it. One reason is that observers, being human, have a very difficult time dismissing their previous experience. For example, one experienced observer correctly noted that page 3 delineates the tangram booklet into two kinds of cognitive demands: placing pieces down on outlines versus manipulating and combining pieces in more complex ways. Based on this, she argued that a child's work can be evaluated against the demands of the task; she did not focus on an evaluation against the child's peers: "To me a maybe is determined on that page 3. If you could finish page 3 alright, without any clues, then that equates to a maybe. Because that's where you get stuck."

Relatedly, experience appears necessary to judge the work. As noted earlier, understanding what complexity is in the Pablo® task is partly determined by the observer's experience. An experienced observer, when asked how she's "trained to interpret what you're looking at" replied that it's a "two-sided" enterprise, one entailing using the checklists and the other: "I think a lot of it is background knowledge."

Maintaining the classroom-based reference is also complicated by a fifth scoring category: "wow." "Wow" is reserved for work that is so wonderful that other children's efforts in the classroom shouldn't be penalized by being judged against it. This score is linked to observers' experience in other classrooms:

> ... then there's also the designation of a 'wow.' That seems to go beyond the definitely. Somebody who just blows everybody else out of the water. And, in fact, that's something that seems to be related to the experience of the observers. For example, you know I don't know how many of these things Judy's done. Probably pretty close to a bizillion. You know, so people who have had a lot of experience seem to really know when something that a child has done is just truly amazing and [they] can do the 'wow' designation.

I asked this observer: "Am I right in recalling this, that you're not supposed to look to other classrooms for --?" She replied: "But you can't, but you -- yeah you're right. But I know that I do that [consider performances in other classrooms]. Because that's my experience." Thus, "wow" places the observer in a complex situation: she must reference her previous experience to determine "wow," and dismiss that same experience in order to judge the students' work according to the classroom standards that DISCOVER officially advocates.

In addition to creating a complex situation for evaluation, the classroom-as-its-own reference group has other downsides. For one, establishing gradations for each score within each classroom is very time consuming, especially for storytelling and Pablo®.

A more serious cost of the classroom-based reference group concerns the validity of the assessment for the purpose of identifying gifted youngsters: What does it mean to be gifted if equivalent performances by Navajo fifth graders in one classroom are scored as "probably" strong, while in another classroom the performances would be considered "definitely" strong?

> Observer 1: But her probablies [the students who receive a score of "probably" in this teacher's class] are like, in another circumstance, we'd be tapdancing if we got some of those.
>
> Observer 2: That's probably true. That's right.

While the desire to compare children only to others with the same classroom experience is viewed by DISCOVER's designers as the fairest approach, the identification as "gifted" which flows from this is problematic. Such identification is akin to describing as "fast" the man who escapes an attacking bear not because he is a good runner, but only because his companion opted first to put on shoes (and got eaten). The identification only describes a child relative to some 25 classmates; it tells us little about what a child actually did or can do.

At this point in time, designers, observers, and teachers in the two schools do not express strong confidence that the children selected are markedly able. Rather, as one observer explained, they are are "hopeful" that this is the case and "confident that we are doing it [the evaluation] in good faith." Maker has said that they have collected a lot of data from the youngsters to help them examine the validity of their identification, but "we haven't looked at it yet."

While there are costs to the classroom-based reference group, this approach also has advantages. Maker argues that it treats children as fairly as possible by not comparing them to others who may have had a richer base of experience. Moreover, by looking only at strengths relative to one's peers, the DISCOVER team actually identifies children who then get some form of enriched instruction. These same children would get no such opportunities using traditional identification methods.

To summarize, the analysis of Condition 4 reveals that DISCOVER's scoring procedures are not as clear as they should be. In the storytelling task, this could be improved by more explicitly drawing on criteria associated with storytelling. For all the tasks, a number of little used characteristics that appear on the Checklist could be eliminated.

One way to ameliorate problems generated by the classroom-based standard, and yet still remain fair to the children, is to establish local norms above the classroom level, perhaps for each of the nine LEAs DISCOVER works with. Left as is, the validity of DISCOVER's work will remain problematic. Children identified by DISCOVER are "escaping the bear." It is largely unknown whether they are, or will turn out to be, gifted. Though the same holds true for children who are identified via traditional methods -- they rarely turn out to be gifted adults (see Chapter 1) -- traditional psychometric efforts are not threatened by such facts. DISCOVER, as a new and underfunded enterprise, must struggle with it.

Condition 5: Observer Reliability

As Cronbach (1989) Messick (1989) and Shepard (1993) have all noted, the validity of an assessment should be built on arguments from a variety of sources that a particular assessment is valid for a particular use. One way the DISCOVER team appears to be building an argument for validity is via preliminary studies of observer reliability.

When asked what gave her confidence that DISCOVER was actually identifying youngsters who are gifted, Nielson noted that "our inter-rater reliability is really very high, especially with experienced observers, which leads us to believe that people are seeing very similar things in the same children." All three of the assessment designers I interviewed referred me to studies supporting the reliability of the DISCOVER work that were conducted by Sarah Griffiths, a graduate student of Maker's.

In one of these studies, Griffiths (n.d.) found a moderate to high degree of reliability between an "original," observer who conducted live, on-site observations, and two trained observers who independently scored videotapes of 25, 9-13 year old Navajo students on the Pablo® tasks. With regard to assigning the score of "definitely, the original observer and Griffiths agreed 100 percent of the time; the original observer and a second, off-site observer agreed 75 percent of the time; and Griffiths and the second off-site observer agreed 84 percent of the time. (Reliabilities for scores below definitely were lower). Griffiths asserted that "observer experience is a likely factor" in lower reliabilities attained between the second, off-site observer and the other two observers. Both she and the original observer had participated in over 30 observations. The second, off-site observer had participated 12 times.

In a second study, Griffiths (n.d.) compared the inter-observer reliabilities of two novice observers (less than 15 observations), two experienced observers (between 16-30 observations) and three expert observers (30+ observations). The reliabilities were based on live observations of 91 students between the ages of five and eleven. Again, she found the highest reliabilities in the definitely category and that reliabilities increased with observer experience. She also noted that training would help to increase reliability (Griffiths, n.d., p. 30).

According to Griffiths (n.d.), these studies suggest DISCOVER has achieved inter-observer reliability. However, in combination with the analysis of observer training presented above, Griffiths' work indicates that the reliabilities DISCOVER obtains in the field are likely much lower because observers on the team are often "novice level." For example, of the six observers who participated in the teams I observed, three (including myself) had fewer than eight previous observations. One was the site-based teacher with intermittent experience whom Rogers believed needed additional training. Only two were experts. Thus, Griffiths' work highlights that inter-observer reliability is certainly possible for DISCOVER, but that it is probably not present in actual practice.

Despite these issues, DISCOVER may still be more reliable than traditional tests in identifying Navajo students. Brenda Romanoff, another of Maker's doctoral students and a teacher of the gifted in Charlotte-Mecklenburg, has compared DISCOVER and Ravens for their consistency in identifying the same group of 61 Navajo elementary students as gifted over a four-year period. Romanoff (n.d.) found that the DISCOVER process was more likely than Ravens, to identify the same students year to year. The Raven's identified a total of 32 children at some point over the four years. However only four of these were consistently identified across all four years. For DISCOVER, 57 were identified at some point during the four years. Of these 31 were identified across all four years. Romanoff suggests that "since a large number of students are consistently identified using the DISCOVER assessment, the procedures associated with DISCOVER are more effective in identifying the strengths of children from the targeted minority group..." (Romanoff, n.d., p. 17).

It is possible that Romanoff's findings may be affected by other variables. For example, some of the same observers are observing the children year to year and may bring to subsequent assessments favorable biases about some children. Even if if this is not so, and even if DISCOVER is more consistent than the Ravens, its inter-observer reliability can be made stronger by relying on more trained and experienced observers.

MI-Specific Conditions

The previous five general conditions are the basis for making sound inferences about students' abilities from any assessment. (See Chapter 1.) Such inferences are needed to associate enhanced equity in identification of gifted youngsters with the DISCOVER assessment. In contrast, the three conditions below are aimed at understanding whether the assessment can be associated with MI theory. For MI to have influenced the assessment, these three conditions should be met.

Condition 6: Assesses Abilities Beyond the Boundaries of Traditional Tests

One of the seeming contradictions about DISCOVER is that it is consistently described as drawing on MI (e.g., Griffiths, n.d.; Griffiths & Rogers, n.d.; Maker, 1992, 1994; Romanoff, n.d.; U.S. Department of Education, 1994) even though its identification efforts center on linguistic, logical-mathematical, and spatial intelligences -- the same abilities typically measured by traditional intelligence tests.

There are at least two reasons why DISCOVER designers nevertheless link the theory with their efforts. One, discussed in more detail in the next section, is that their efforts are "intelligence-fair." A second reason may be that observers record information pertinent to a wide range of intelligences. For example, the Problem-Solving Behaviors Checklist includes a series of characteristics under headings for the seven originally proposed intelligences, except music. Thus, a broad range of intelligences is woven into the assessment. Yet, the actual identification of giftedness is not made on this broad range. It is focused on two language tasks (storywriting and storytelling), two tasks entailing spatial problem solving (Pablo® and tangrams), and one or two tasks of logical-mathematical strength (the math sheet; tangrams taps this to some degree). Two definitelies out of these five tasks yields identification.

Contextual issues to be examined in the last chapter reveal why DISCOVER does not draw on a wider range of intelligences. To preview briefly one issue, school systems value strengths in language and mathematics most. Therefore, they are less likely to select an assessment that seeks out a wide range of strengths. As Maker explained:

> First of all, you have to get people to believe that musical and bodily-kinesthetic would be important to assess. Because they don't see their task as having anything to do with development of bodily-kinesthetic and musical intelligence. We're only now getting some schools that want assessments in those areas. Because they now see the importance. And so, if you're going to develop an assessment, you start where you think somebody's going to use it.

For this and other reasons, DISCOVER does not meet the condition of assessing abilities beyond the boundaries of traditional tests.

Condition 7: Intelligence-Fair

One of the reasons that DISCOVER designers feel their assesssment draws on MI is because it is intelligence-fair. That is, DISCOVER assessments allow children to be identified on the basis of performances in a range of media relevant to different intelligences. DISCOVER's designers have drawn on Gardner's notion of first- and second-order knowledge (Gardner, 1991a, 1991b) to help them distinguish between their two traditional assessments and the three alternative tasks. First-order knowledge does not engage the formal, typically written representation system that is central to schooling. Second-order knowledge entails being able to express first-order knowledge in a formal system of representation. Thus, storytelling taps first-order knowledge, while storywriting requires second-order knowledge.

One of the useful things about this scheme is that it allows strengths to be seen (and justified) in forms other than those typically recognized by school. As Nielson put it:

> We think that our Pablo®, and our tangrams, and our storytelling activities call on first-order knowledge, whereas our storywriting activity and our math worksheet call on second-order knowledge. And we think that first-order knowledge has been very strongly neglected in the past.

Another way that DISCOVER is intelligence-fair is that it seeks to diminish the language demands of the assessment. The team keeps its directions short and simple, as noted earlier, partly for this reason. As Rogers explained: "We've tried very hard to not make language be a barrier for any child's abilities -- in other words, not let language get in the way of allowing children to show us what they're capable of doing."

Rogers cautions fellow assessors to go beyond impressions left by a child's language usage in order to evaluate the abilities actually under consideration. She tells them that in Pablo®, "I don't mind if you talk about the language that the kids used. But that's not really the important part of this task. And please don't allow that to interfere with your decision in this task."

Alongside lessening language demands, DISCOVER's alternative assessments employ engaging, hands-on materials that allow children to demonstrate their abilities in intelligence-fair ways. Maker sees the hands-on approach as central to uncovering children's strengths:

> [A] lot of tests are not engaging for children from a background in which you do hands-on stuff: you work with animals, you get outside, you work in the earth, you make jewelry, you make stuff, you do things with your hands.... And how motivating is it to fill in a bunch of bubble sheets? I mean bubble in stuff on a standardized test? I don't think that's motivating to them. Other than those [kids] whose family say, 'You know, you've got to do well on this thing. So bubble it in.'

To be intelligence-fair, children are not evaluated via 'bubbles.' Instead, they are given Pablo® tasks and tangrams that directly tap spatial problem solving. Rather than being handed a pencil to demonstrate language skills, children are given toys to encourage storytelling. It would be wonderful if logical-mathematical abilities could be assessed somewhat more by such intelligence-fair methods (though tangrams partly draws on this ability). Still, overall, DISCOVER succeeds in being intelligence-fair.

<u>Condition 8: Domain-Based</u>

Gardner asserts that an intelligence is an ability to make products or solve problems that are valued in one or more cultures (Gardner, 1983). In other words, intelligence is manifested in practices or "domains" valued by a culture. Thus, MI implies -- and Gardner's later writings (e.g., Gardner, 1991a) explicitly state -- that assessment needs to be embedded in practices recognizable by the surrounding culture. (See Chapter 1.) Thus, an assessment of language should employ culturally valued uses of language, such as storytelling or essay writing. Conversely, an assessment of language that entailed inferring the meaning of nonsense words placed in sentences would not be considered domain-based (except in a culture of test designers!).

In her writings, Maker supports the view that culture shapes the expression of abilities and that MI provides "a helpful way to examine giftedness across and within cultures" (Maker, 1993). However, only the DISCOVER tasks that assess linguistic ability are clearly embedded in a domain: Storytelling is intrinsic to virtually all cultures. The storywriting task fuses this culturally valued activity with the valued, "second-order" notations that represent it. The math worksheet is a traditional school activity, one that may draw on practices valued in school but not practices generally modelled or used in the wider culture. The Pablo® task is culture free; tangrams might be a domain-based task in China (where this activity originated) but not among the Navajo or in most other communities in the United States.

The reasoning behind using Pablo® and tangrams resembles the reasoning behind standardized tests that ask children to uncover the meaning of nonsense words: Tasks such as these are novel. Thus, they supposedly allow children's underlying abilities to be manifested while controlling for differences in children's experiences. The idea of "controlling for" experience by a variety of techniques, including standardized instructions and novel tasks, is at the heart of almost every psychometric test. There are benefits in using this approach. One is that it doesn't totally sever DISCOVER III's work from the psychometric mainstream. One cost is that such tasks do not come with any meaningful criteria for judgment: What makes for a good Pablo construction is much harder to ascertain than what makes for a good yaahbichii, a product of Navajo culture for which standards exist. Another cost of domain-free tasks is that they attentuate DISCOVER's tie with the theory that is supposed to undergird it.

CONCLUSION

The analysis of DISCOVER III reveals that it does not meet the five general conditions that are needed to make inferences about students' performances from any assessment. Therefore, it is not yet reasonable to associate changed outcomes with the assessment. In the future, this association may be established partly by drawing on a well-trained and experienced observer team. This, in turn, will help DISCOVER achieve inter-observer reliability. In addition, such trained observers will need clearer scoring procedures to draw upon.

The analysis of DISCOVER against the MI-specific conditions indicates that it is also not possible to associate the assessment or increased identification with MI theory. Clearly, however, DISCOVER has made strong use of intelligence-fair practices suggested by Gardner (1991a). In addition, both the description and analysis of DISCOVER's efforts have pointed to a number of other strengths. First, the children typically understand what it is they are being asked to do. Second, the materials and procedures support children's best efforts. They engage children, and they enable them to be identified without having undue dependence on language or notational skills.

Maker and her colleagues are also due enormous credit because they have undertaken difficult, pioneering work. There were few historical precedents for using MI to identify underserved youngsters, nor was there a framework or set of conditions to support the construction of their assessment. My hope is that this framework of eight conditions makes it easier to discern how to strengthen the DISCOVER assessment and make it less vulnerable to critics and funding shortfalls.

Finally, it is important to emphasize that DISCOVER's work has been influential. Schools and districts in this country and beyond have adopted the DISCOVER process. As we will see in the next chapter, others have also modified and adapted it. The influence DISCOVER wields is partly due to the spirit of the team's members -- notably to its emphasis on finding children's strengths -- as well as to their energy in mapping out challenging, new territory.

1. The Raven's Colored Progressive Matrices is one version of the Raven's Progressive Matrices test. There is also the Raven's Standard Progressive Matrices, a version commonly used for elementary students and young adults.

2. In the Navajo language, there is no word for "gifted," according to Susan Bartley, the director of All Can Excel, the enrichment program for the Chinle Unified School District.

3. In order to focus their investigations, the designers of DISCOVER III decided not to administer the assessment school wide in 1995-1996.

4. I had hoped to include examples of the instruments, but I was not given permission to do so. Maker and her colleagues have copyrighted these documents and plan to produce a manual for the complete assessment process.

5. Nielson, commenting on a draft of this chapter for both Maker and herself, asserted that the storytelling task could involve the creation or retelling of a story: "Students are not told they have to create a new story; they are asked to tell a story using some or all of their toys. If a student chooses to retell an existing story, the focus should be on the act of storytelling...." (personal communication, February 18, 1997, p. 4). That the observers had difficulty knowing whether it was acceptable for a child to retell an existing story, and spent considerable time debating this point, speaks to the need for more and clearer training of the observers.

Chapter 3
CHARLOTTE-MECKLENBURG SCHOOLS: HOMESTEADERS

THEORETICAL BASES FOR THE PROBLEM SOLVING ASSESSMENT

The staff of Charlotte-Mecklenburg Schools' Program for the Gifted evolved two different assessments said to draw on MI. The first, Project S.T.A.R.T. (Support To Affirm Rising Talents), was a collaboration between the Program for the Gifted ("PG") and the University of Virginia. It was funded by Javits from October 1992 until October 1995. S.T.A.R.T. drew on both Spectrum (see Chapter 1) and DISCOVER assessments in an effort to identify promising, poor and minority K-1 students. Once identified, youngsters were provided enriched classroom environments, mentors, and family outreach services.[1]

The second effort said to be influenced by MI is the Problem Solving Assessment ("PSA"). While S.T.A.R.T. assessments identified youngsters "at promise" and provided them with enriched environments, the PSA is the primary tool for identifying youngsters for gifted services in Charlotte. Because of its central role, and because S.T.A.R.T. funding had ended by the time of my visit, I have focused this chapter on the PSA.

Though it was not directly funded by the Javits Program, the PSA evolved partly with Javits' resources. Mindy Passe, the coordinator of Project S.T.A.R.T., worked closely over several years with Brenda Romanoff and other PG staff members to develop the PSA. Passe noted that a number of consultants who came to Charlotte, including Maker, were supported with Javits funds. The PSA can also be traced to Javits in that a major initial influence upon its format is Maker and her Javits-funded DISCOVER. The

PSA draws on the problem solving continuum, incorporates some of DISCOVER's activities, and emphasizes spatial, linguistic, and logical-mathematical intelligences. Maker's influence is evident in the definition of gifted youngsters used in the Charlotte-Mecklenburg Schools ("CMS"):

> Gifted students are those who demonstrate extraordinary problem solving abilities in the linguistic, logical-mathematical, and spatial intelligences. When presented with an open-ended or challenging problem, extraordinary problem-solvers demonstrate creativity, critical thinking, and task commitment in order to reach a productive solution[2] (CMS, 1994b, p. 3).

Although the theoretical models of the PSA are traceable to Maker's adaptions of Gardner's ideas, the teachers and administrators in Charlotte's Program for the Gifted have settled the territory DISCOVER first explored. The Charlotte designers, like homesteaders, have sought to adapt the terrain to their own particular needs.

HISTORICAL AND CULTURAL CONTEXT

"If you always do what you've done, you always get what you got."

- An old Southern saw, according to Brenda Romanoff, PG teacher and a developer of the PSA

Charlotte is located in the south-central part of North Carolina. The city and surrounding Mecklenburg County is home to 597,000 people. For the past several years, Charlotte-Mecklenburg has undergone an economic boom. A number of corporations from around the country have relocated there. It is a center of banking, finance, trucking, and wholesale distribution. Unemployment is low -- approximately 3 percent -- and the population has grown at 2.6 percent per year since 1990 (Stewart, personal communication, 1996.)

Charlotte-Mecklenburg Schools comprise a large, county-wide district encompassing 550 square miles and 123 schools. In 1994-95, the district had over 88,000 students and 10,000 employees. The district is described as urban (Charlotte Observer, 1995), though visits to various schools in the county reveal inner city, suburban, and even rural areas. Approximately 40 percent of the district's students are African American, and approximately 60 percent are white.

Issues of race run deep in the Charlotte-Mecklenburg Schools. With regard to integration, North Carolina appeared politically moderate relative to other southern states. Yet, in 1964, ten years after the Supreme Court ruled in *Brown v. Board of Education* that segregated education was unconstitutional, only three percent of the state's African American students were assigned to "white" schools (Douglas, 1995). After extensive and often bitter legal battles among Charlotte's School Board, the NAACP, and other community agencies, the Supreme Court's 1971 decision in *Swann v. Charlotte-Mecklenburg* upheld a ruling supporting the most extensive court-ordered busing in any U.S. city. The ruling touched off national debates and protests (Douglas, 1995).

In 1974, after much legal maneuvering, the courts, aided by a consortium of community groups, put into place an extensive busing plan that created a highly integrated and reasonably fair system. Into the mid- to late-1980s, Charlotte remained committed to school integration (Douglas, 1995; Morantz, 1996) and "Charlotte's resolution of the busing issue ... was a source of local pride..." (Douglas, 1995, p. 251). More recent efforts to minimize busing by former Superintendent John Murphy[3] and a business elite new to Charlotte appeared to have increased segregation in school.

However, further segregation was defeated in 1995 when a new school board that largely favoring integrated education was elected (Morantz, 1996).

Anne Udall, who was hired to head Charlotte's Program for the Gifted late in 1992,[4] claimed that *Swann* is "a real important piece of the history around here" and continues to have an impact on people's thinking. Despite this, programs for the gifted remained largely the domain of white students until the early 1990s. According to one staffer, gifted education has been widely used as "a white track." Charlotte's gifted program had been an "elitist, isolated, white-only program," a pattern only now "beginning to change."

The underrepresentation of African American students in Charlotte's gifted education programs was partly due to traditional identification methods. For children to be identified in Charlotte, they first needed to be referred for testing. However, as noted in Chapter 1, African American students are markedly under-referred. One PG staff member asserted that "...a lot of the [teacher] referral patterns indicate a very unspoken bias." In Charlotte, not only teachers, but other adults could refer youngsters for assessment. However, Udall noted this fact was not always widely understood or shared.

Once referred for assessment, actual identification practices were governed by state policy. For a district to receive funds for a gifted student, that student must accumulate 98 points from three sources: up to 50 points (for a score in the 99th percentile) from achievement tests in reading and math. Another 50 (again from a score in the 99th percentile) can accrue from performance on an IQ test. Up to ten points -- amounting to "a few bonus points" -- can come from a student's grades. Given the reliance on teacher referral and standardized tests, it is not surprising that only a small percent of Charlotte's identified students were African American.[5]

To increase minority participation, for several years Charlotte designated students who had between 88 and 98 points as "academically talented" versus "academically gifted." Such students received gifted services identical to those who scored at the state-required 98 points. Nevertheless, under this policy only between 8 and 12 percent of the district's identified youngsters were African American.

The traditional approach to identification began to give way when John Murphy became superintendent in 1991. Soon after his arrival, Murphy held "town meetings" across the county to hear what citizens had to say about the schools. A number of concerns were voiced about gifted education: Parents felt there wasn't "enough of it, especially at the elementary level." The interpretation of this is not transparent: the problem was *not* that the gifted program excluded minorities. Instead, the program was not having much of an impact on the youngsters already in it. In addition, parents felt that the gifted program had been adrift. For two years, it had been without a leader.

These concerns led Murphy to tell the teachers of the gifted that "we've got to do something." The shape that "something" first took was the appointment of a task force comprised of seven teachers in the elementary gifted program (four of whom were interviewed for this chapter), principals from various levels of schooling, a central office administrator, and representatives of Charlotte's PAGE (Parents for the Advancement of Gifted Education). The committee was co-chaired by Carol Reid, a teacher of the gifted, and now the PG Program Specialist, and by Professor Carolyn Callahan, of the University of Virginia.

When the task force convened in the fall of 1991, its actual charge seemed to be to enhance service delivery at the elementary level. However, Murphy gave the task force leeway to investigate gifted programs throughout the United States and to devise a system that made sense for Charlotte-Mecklenburg. Mindy Passe, a member of the task force, said "The diversity issue -- that was an important piece for Murphy.... The task force was a way of handling both issues: parents' concerns [with elementary level gifted education] and equity."

Among the goals the task force set for itself was to examine and address the identification process. Reid noted that task force members wanted an assessment that was aligned both "with current thinking about intelligence," and with "the service delivery model" or gifted curriculum. The task force also sought an assessment that was better able to detect the gifts of underrepresented youngsters.

The task force members undertook a lot of reading and reflecting. They were especially attracted to both Gardner's and Sternberg's ideas. (See Chapter 1.) However, according to Passe, Sternberg's ideas were not being implemented much in schools, so there were few models to consider.[6] In addition, Reid noted, that relative to Sternberg's ideas, Gardner's are accessible.

During 1992, "the year of the consultant," the task force met with several people actually applying Gardner's ideas to assessment. Among these were DISCOVER's Maker and Rogers, Waveline Starnes from Montgomery County (whose work is discussed in Chapter 4), staff of Brooklyn's Javits 7+, and researchers from Project Zero who had helped to develop Spectrum assessments. (See Chapter 1.) From these various readings and meetings, the task force began to revise Charlotte's educational offerings, crafted its Javits grant application for S.T.A.R.T., and started to forge a new identification system.

Because of state policy, along with the new PSA assessment, students can still take the battery of standardized tests and be admitted under the state's 98-point system. This, according to Udall, "works in everybody's favor, because it gives parents an alternative to entering the program, and it gives us kind of a safety valve."

GIFTED EDUCATION IN CHARLOTTE-MECKLENBURG

Gifted education is a much more complicated and differentiated enterprise in Charlotte than it is in Chinle. (See Chapter 2.) At the elementary level, gifted education in CMS includes a variety of programs. The most prominent is "Encounter," a pull-out enrichment program taught by teachers trained in gifted education. It offers fast-paced, small group work to identified children for approximately six hours a month. It seeks to "make students aware of the connections in all knowledge," to develop critical thinking skills, leadership, and teamwork and instill a "sense of community." Another program is Catalyst. This calls upon PG teachers to work with classroom teachers to develop enriched curriculum for more classroom-based instruction for the gifted. In addition, the district opened three elementary gifted magnet schools in 1993-1994. At third through fifth grade, these magnets serve only identified youngsters. In kindergarten through second grade, these schools are "learning immersion" sites aimed at enhancing the numbers of identified minority students largely from the neighborhoods around the schools.

INQUIRY INTO THE PROBLEM-SOLVING ASSESSMENT

PREASSESSMENT

The new identification system in Charlotte has two phases. The first is preassessment in which youngsters are exposed to problem solving akin to that demanded by the PSA. During this first phase, PG teachers provide weekly preassessment lessons to each second grade classroom until the actual PSA is given. There are a minimum of three separate preassessment lessons for each second grade classroom. During these lessons youngsters are given instruction and activities that focus on linguistic, logical-mathematical, and spatial intelligences. Preassessment lessons may include a wide variety of materials, such as Lego® blocks, maps, and tesselation puzzles for spatial intelligence; pentominoes, and pattern blocks, and story problems for math; and word- or speaking games like Scrabble for linguistic intelligence.

During the preassessment lessons, both the PG who serves the school and the regular classroom teacher observe and take notes on what the individual children in the class are doing. Some of this work is also collected into a child's Second Grade Classroom Portfolio. This is a preprinted manilla file, which provides information and activities associated with the three intelligences that are assessed. A teacher checklist on the back of the portfolio highlights behaviors associated with each intelligence. For example, for linguistic intelligence, characteristics include "is an avid reader" and "enjoys telling detailed and expressive stories." The teachers check these behaviors along a four-point scale: not evident (which means rarely, if ever, evident), evident, strongly evident, and always evident. (See Appendix F.) The latter is akin to DISCOVER's "wow" (See Chapter 2) and designates extraordinary performance. It is rarely used.

Formal identification in CMS is a high stakes affair. It allows those in the learning immersion programs, as well as others across the district the possibility of attending the gifted magnet program in third through fifth grades. According to Udall "the biggest reason" parents want identification at elementary school is that it allows "entrance in the middle school gifted classes." This fosters preparation for the International Baccalaureate or other demanding high school curricula. Identification is also a formal requirement for enrolling in high school AP classes.

As in many public school systems, the distinction between high level and general courses even in the same high school is marked and carries genuine consequences. To illustrate, in a single high school in Charlotte, I attended a 9th grade general English class, in which mostly African American students were cutting pictures from popular magazines to accompany a worksheet about different emotions. In a 10th grade International Baccalaureate class, in which there was one African American student, the students were discussing Hannah Arendt and the tensions between the *viva activa* and *viva contemplativa*. There was almost no possibility of students from the 9th grade general English class going on to participate successfully in the 10th grade IB class. They had not been prepared via earlier curriculum challenges to engage in such a high-level discussions.

As these examples reveal, in Charlotte and elsewhere, early participation in gifted education, or lack thereof, ultimately contributes to divergent educational experiences. Identification helps set youngsters on a path that is far more likely to prepare them for admission into selective colleges.

After all the preassessment lessons are completed for a given class, the PG and the classroom teacher together review each child's second grade portfolio. The teacher can also bring in other information based on her own observations and work with the students. Based on their review and discussion, they jointly refer youngsters to be assessed for the actual PSA.

THE PSA

Once the preassessment phase is completed, referred children take the PSA. The PSA is administered in each of the elementary schools. Different versions of the PSA have been developed for each grade from second through fifth. I observed the assessment of second graders in two schools in late October 1995. The observations of the debriefing session included efforts to evaluate all parts of the assessment. The observation of the assessment administration did not include the tangrams and Pablo®. However, interview and documentary data reveal that CMS' Pablo® and tangram tasks are very similar to those used by DISCOVER III.

The two schools I visited differed in a number of ways. Berryhill School is located in the western part of the county, in an area bordering the airport. The surrounding area looks sparsely settled, with very modest wooden houses and occasional small storage buildings that appear to serve the airport. Several of the observers commented that the school is hard to find, heightening, for me, a sense of its remoteness.

The school itself is a one-story, white concrete structure. It was built about 1970 originally as an open school. Later, it was divided into more-or-less regular classrooms. Despite somewhat haphazard architecture, the school environment felt warm and welcoming. From behind the gym door on the right side of the wide entryway, the sound of a small group, woodwind lesson filtered in. Along the corridor walls were youngsters' pumpkin pictures made of crushed and glued orange crepe paper. There were also large posters by children for a campaign against drugs and violence on television, an initiative underway across the district that was getting news coverage on local TV.

At the time of my visit Berryhill had 454 students, kindergarten through Grade 6. The student population is 45.8 percent white, 51.3 percent African American. The poverty rate is said to be high across both African American and white children. Almost 65 percent of the students are on free or reduced lunch, 69 percent come from homes with incomes below $25,000 per year, 47 percent live with both parents. About 34 percent of the students' mothers had some college or technical school education (Charlotte Observer, 1995). Because the district in 1995-1996 sought to cast as wide a net as possible, all second graders from this and the other "Challenge Team" schools -- some 10 schools with high poverty and few identified gifted youngsters -- were assessed using the PSA.[7]

The second school I visited was the McKee Road School. It is located in the southeast part of the county, in an area quickly developing from farmland to suburb. On the way to the school were several stands of new townhomes, but on the road leading right up to the building there was still a working farm.

The school itself was built about 1989. Though it is a large single-story white structure, made to feel even larger by many big windows, it has been outgrown by a rapidly expanding student population. The school was built for 750 students. At the time I visited, it was serving about 930 K-3 students. McKee students' circumstances are quite different from those at Berryhill: About 87 percent are white, 11.2 percent African American. Ninety percent live with both parents. Just over 90 percent of students' mothers have had some college or technical school education. Only 9.9 percent get free or reduced lunch, 6.7 percent come from families with incomes lower than $25,000 (Charlotte Observer, 1995).

The PSA was a qualitatively different endeavor at the two schools. At McKee Road, the youngsters were quite aware that they are being screened for identification for the Program for the Gifted. It is, in essence, a high stakes test for them. Thus, the children were serious and quiet. In contrast, at Berryhill, the children are not junior SAT takers. They appear largely not to know or not to understand that various educational opportunities rest on their performance during the PSA. As a result, during the PSA, they seemed like regular second graders: animated, occasionally distracted or confused, but by and large engaged in activities that are different from their everyday school experience.

CHANGES IN IDENTIFICATION RATES ACCOMPANYING THE PSA

As noted above, prior to using the PSA, between 8 and 12 percent of the identified students for gifted education were African American. Nearly all the rest were white. A study based on a random sample of 600 student files found that with the 1994-1995 version of the PSA, about the same proportion of the district overall is identified. (Between 10 and 12 percent of the district's students are usually identified). The percent of female and male students identified remained nearly equal. However, the percentage of identified minority students roughly doubled, to 19 percent of those identified (Reid, Udall, Romanoff, & Algozzine, in press). In 1995-1996, the year in which my observations took place, the identification rate for minority youth was 18 percent.

Yet, across the county's schools, identification rates vary widely between approximately 3 and 34 percent of the school population (Charlotte Observer, 1995).[8]

The differences in identification rates are reflected in the two schools I visited. At Berryhill, five children, or between five and six percent of the second graders were identified. This percentage remained constant over two years (1993-1994, 1994-1995). According to Becky Workman, the school's PG teacher and a former task force member, Berryhill has one of the highest identification rates among the county's "low pop" schools. In the past, only two or three students in the school's *entire* K-6 population had been identified.

However, identification rates for African American youngsters at Berryhill remain low. Of the five children identified among the second graders during my visit, one was a Native American, the rest were white. At McKee Road in 1995-1996, 79 children, representing 31 percent of the second graders, were identified. "Two or three" minority youngsters were identified among this group in 1995-1996. In 1994-1995, approximately 30 percent of the second graders were identified, and the number of identified minority students was again very small, according to Steve Houser, the school's PG teacher. Clearly the identified minority population is disproportionately low for a school with 11 percent African American students. It is also low for Berryhill, where more than half the youngsters are African American.[9]

While the two schools in this investigation do not reflect the sorts of changes found across the district, they do provide the observational data to help illuminate whether such district-wide changes can reasonably be attributed to the PSA. The observations highlight the tasks that are used, how the tasks are administered, and how information gathered from students' performances on the PSA is evaluated. After describing the assessment along these dimensions, I analyze it against the general and MI-specific conditions introduced in Chapter 1.

DESCRIPTION OF THE PSA TASKS AND PROCEDURES

The Problem Solving Assessment consists of nine activities, most of which include several tasks. Seven of the activities are administered on a single day by an assessment team that visits the school. Two activities are administered by the school's PG teacher prior to the team's visit.

During the activities with the assessment team, four or five children work at a single table or cluster of desks with one observer. Typically after an activity or group of activities related to a particular intelligence, the observers rotate, so that each child is observed by several different adults. The assessment team's activities begin in the morning and continue after lunch, for a total of about four hours.

The assessment activities fall along a continuum from traditional, standardized, paper-and-pencil tests to more alternative activities. At the traditional end is the standardized Matrix Analogies Test. The alternative activities include those borrowed and adapted from DISCOVER: Pablo®, tangrams, and storytelling. In order to impart a sense of how the assessment proceeds for the student, I am describing the tasks in the order in which they are administered.

TASKS ADMINISTERED BY THE PG TEACHER

The Story Writing Task

For children referred to take the PSA, like those at McKee, the storywriting task is administered by the school's PG teacher to groups of approximately of five to eight children outside the students' regular classroom. At Berryhill, a Challenge Team school in 1995-1996, whole classes of second graders took the storywriting task.

The students are given a "story starter" to prompt their storywriting. In 1995-1996, the story starter asked the children to write about the place they would go if they could spend a day anywhere in Charlotte. Children typically have 30 minutes for this task, but they can stay with the PG teacher and work longer if they wish. (They cannot take the assignment home with them.)

Children's written stories appeared quite variable. At Berryhill, some children wrote basically no story.

A highly-rated story for a second grader at McKee was:

> One morning I was getting ready for school. I was listening to my TV. I looked at my TV. Then I started to put clothes on. I looked at the TV again. It said, 'Today's a special day.' I said to myself, 'What could be so special?' I looked again. This time it said, 'You can go anywhere you want.' I jumped for joy. I went downstairs to ask my mom if she saw it on TV. I asked her. She said she saw it. She said I could go anywhere in Charlotte. I got on my bike and pedalled all the way to Zones. There I played games and won prizes. Then at 9 o'clock, I went to see the Hornets play against the Bullets. The Hornets won. After the game, I went home and got in bed.

<u>The Matrix Analogies Test</u>

The Matrix Analogies Test, or MAT, is a standardized measure drawing largely on figural reasoning like the Raven's (See Chapter 2). Except for children in Challenge Team schools, children also take this test in small groups, usually five students at a time outside of their regular classroom. The MAT is administered within a few days of the story writing task. The MAT is a timed test, allowing 25 minutes for 35 questions.

Students' Experience of the Activities Administered by the PG Teacher:

As in Arizona, when children are assigned tasks in situations that appear test-like, they behave in a test-like fashion. They work on their own. They do not talk. They usually engage the task with quiet concentration.

PG Teachers' Role During the Tasks

During these two tasks the PG teachers are responsible for reading the directions and ensuring that children understand what they are supposed to do. On the MAT, this includes guiding the youngsters through an example and showing them how to fill in a bubbled answer sheet. For both tasks, the PG teacher also ensures that the students work on their own. At the same time, she attempts to create a comfortable atmosphere by establishing some rapport with the children. In addition, for the storywriting task the PG teacher records the students' product and process behaviors on the "Problem-Solving Behavior Observation Card," more manageably known as the "yellow card." This is a 6-sheet instrument, printed landscape fashion on yellow, 11x8.5 paper. It folds-out menu style, with a page or half-page devoted to each section of the PSA, except the MAT. (See Appendix G.)

Scoring/Evaluation of the MAT and Storywriting Task

In 1995-1996, the MAT was scored either by the school's PG teacher or forwarded by the PG teacher to be scored by the assessment team. In 1996-1997, all the MATs are forwarded to the assessment team for scoring. The scoring entails totalling the correct answers and assigning a stanine score. The storywriting task is scored by the assessment team on the four-point scale of "not-evident" to "always evident," described above.

Further details pertaining to scoring will be discussed after the remaining PSA tasks are described.

TASKS ADMINISTERED BY THE ASSESSMENT TEAM

The Storytelling Task

In 1995-1996, the first task administered to the students by the assessment team was the storytelling task. The task, and most of those administered by the team, begins with instruction that highlights the salient aspects of what is to be done. The instructions for all the tasks are clearly spelled out for the observers in a manual prepared by the Program for the Gifted (Udall, Reid, & Romanoff, 1995). These instructions were very closely followed by the observers at McKee Road and Berryhill.

The observers at each table ask the children to think about stories that have been told to them and then they ask, "What makes a good story?" After the children suggest various elements, the observer mentions those left out, such as place, action, or detailed descriptive words. The observer encourages the students to think about interesting topics, and she highlights some nuances of storytelling, like "Using your voice to show feelings" and "using your body to show gestures or action." The observer underscores that, "The most important part of the storytelling is the words you choose to use" (Udall, Reid, & Romanoff, 1995, p. 14).

Then the children are told that they are going to tell a story to the observer. The story should be one they make up, not one that they've heard or seen before. To help get them ready, the observer leads the children through a visualization exercise: "Imagine you are at your favorite place. Look around and notice what is there. Notice the colors. Notice the smells.... "

After this, the observer asks the children to choose a small plastic animal from among those she places on the table: a panda bear, a camel, an elephant, a raccoon, a rhinocerus, a zebra, a polar bear, a jaguar, and two or three others. They are then told, "You are to make up a story to explain how the animal got to be the way it is." The observer instructs them to provide a detailed story, not a short explanation. The observer tells them their story should last no more than five minutes. When a child is ready to tell his or her story, the observer is supposed "to record the story in a quiet place where there are few distractions and where the child feels comfortable telling his/her story" (Udall, Reid, & Romanoff, 1995, p. 15).

The Students' Experience of the Storytelling Task

The students' experience in storytelling was somewhat different at the two schools. In Berryhill, the children seemed more animated: For example, they moved their toys around, and sometimes smiled, as they told their stories. Children who were not away from the table to tell their stories to the observer were encouraged to "draw a picture of the animal." Perhaps because the CMS storytelling task does not encourage children to play with the toys on the floor or elsewhere, as children in DISCOVER's storytelling are, the room as a whole was busy, without at all being frenzied.

At McKee Road, the students I observed were quite a bit more somber. They waited quietly and politely until it was their turn. They sat still, most barely moving their toy, even during their own storytelling. Part of this seriousness may be due to the fact that the children were not always moved away from the table to tell their story. Thus, they may have felt somewhat inhibited speaking before not only the observer but also their peers.

The stories produced at both schools tended to be fairly short. One observer reported that a child at Berryhill who had picked a zebra said he "couldn't think of anything to say. And finally he said he had it, and his story was going to be short. And it was. Maybe three sentences."

A story judged to be strong at Berryhill was about an elephant, which the observer recounted as:

> Its nose used to be short. It was all squashed in. He was playing with a snake near the water and leaned over. The alligator bit him on the nose and he held on, and he pulled, and he pulled, and he pulled, and he pulled. When he finished, his nose was long, and it's been long ever since.

At McKee, the following story was judged strong:

> All the panthers were white at the beginning of time. One night it was so dark, and the panthers went in the cave. And it was darker than it had ever been. And the darkness went into its skin. But he didn't know it. And the next day, he came out of the cave, and he met an elephant. The elephant said, 'What happened to you?!" And he said, "Well, nothing! "I'm fine." And the elephant said, "Well, no. You're black." And so he goes back to the cave and looks in his crystal mirror. And he realized that he was. And all the other panthers, it was so dark during this period, all the other panthers took the darkness into their fur, and that was why they were black.

<u>The Observers' Role in the Storytelling Task</u>

The observers have a vital role to play in storytelling. They ensure the students understand exactly what is being asked of them. For example, at Berryhill, when a table of children did not offer an example of descriptive words, an observer said that one girl at the table had "a *beautiful* shirt, with purple, pink, and yellow."

Observers must also encourage the students. As Steve Houser, the PG teacher put it: "the assessor needs to be attentive to the student as the story is told. And if they're not attentive, then it's for naught."

In their role as documenters, observers look for traditional story elements as well as unusual or creative narratives. One observer reported "looking for the components of a good story. Good detail, things like that.... [A]t the same time, I'm alert to and recognize something unique."

After listening to each child's story, the observer takes a few minutes to document the child's performance on the "yellow card." For Storytelling and Storywriting, there are some 17 behaviors which an observer has the option to check. There are also two long lines in which to add additional behaviors, and then two printed lines in which to make comments. (See Appendix G.)

The Logical-Mathematical Tasks

The second set of tasks administered by the assessment team consists of several discrete activities aimed at exploring youngsters' logical-mathematical abilities. Each of these activities takes up one or two pages in the Second Grade Assessment Student Booklet, the student's 15-page, yellow test booklet. The set of tasks lasts approximately 45 minutes.

"Part I: Sequences" consists of four, 1- or 2-digit number sequences in which one number of the sequence must be identified by the child (e.g., 12 10 ___ 6 4 2; 80 75 76 71 72 ___.) The written directions in the student test booklet read: "Find the number that completes the pattern." This is followed by a fifth, openended problem in which the students are directed to "make up your own sequence."

Before the students begin, the head of the observer team stands at the front of the room and talks about the concept of patterns. During my observations, the leader, a PG teacher named Ty Fox, asked for examples of patterns. She then drew out the notion that children offered to the sample were inserted into an overhead and discussed before the actual problems were undertaken.

"Part III: Number Logic" follows the format of the preceding sections: four given problems, and an openended question of the same type. Each problem entails two sets of 3- or 4-term arithmetic equations in which one or more terms in each set are replaced by shapes or figures, such as a sun, square, or bell, e.g., $\square + 4 = 6$ AND $5 - \square = 3$. The written directions to the students read: "In these problems, shapes take the place of numbers. In each row, the same shape is always the same number. Write the correct number in each empty shape to solve the problems. Watch your signs" [i.e., plus and minus signs].

Before beginning these problems, Mrs. Fox explained the activity by leading the class through a discussion using three sample questions printed in the Student Booklet (including the one above). They were told to "Think about which number could be placed in each of the squares to make both number sentences true?" Their answers to the sample problems were discussed, before the children began the actual problems.

"Part IV: Fluency and Flexibility" is similar to the final question on the DISCOVER math sheet. (See Chapter 2.) It asks the children to "Write as many problems as you can that have 10 as the answer. You may use the whole page." Before the actual task began, Fox worked through and discussed with the children a sample question that used 2 as the answer.

"Part V: Story Math" is the last activity in this section. In this task, students are presented with three stories, the second contingent upon the first, the third is independent of the others. The stories are each read twice to the children. During the first reading of each story, students are directed just to listen. During the second reading they were to use the paper in front of them "to work this out any way you wish."

During the first story, Mrs. Fox held up 8.5 x 11 inch line drawings of each of the items to be calculated as the story proceeded.

> Your mom has to go to the store. She tells you to take good care of your little brother for a few minutes. Your favorite TV show comes on and you forget all about your brother for a few minutes. Then you hear a crash in the kitchen!! As you walk to the door of the kitchen, you are shocked by the mess. Add up what you see. You see: five Oreo cookies [drawing], a plastic cup [drawing], a jar of peanut butter with a knife inside it [drawing], a quart of milk [drawing], three crackers [drawing], a peach [drawing].

In the second story, the children figure out how much is "left for mom to see" after the child has put away the peach, the cup, and the knife and peanut butter. The third story asks children to keep track of how many baseball cards "Gilbert" has after five days of giving away, trading and buying cards.

Students' Experience of the Math Tasks

In general, the math tasks follow the format of school tests. The children work with paper and pencil, and they are told to cover up the student booklet in which they record their answers. As a result, they worked independently and did not talk to each other. However, they did talk with the observer as necessary, and seemed to feel free to do so, especially at Berryhill. There, for example, they sought and received confirmation that they were doing the tasks correctly. At McKee Road, all the children were clearly engaged in the task. They worked quietly and steadily. Most did not ask for help.

The students' performances on several of the logical-mathematical tasks varied greatly. For example, on number fluency, at Berryhill, students generated between 4 and 20 equations, and at McKee Road, between 6 and 29. At Berryhill, there were several occasions that observers noticed children "had some ability, but they just had no math [skills] to work with" or a child's "numbers maybe are just not sophisticated enough" despite "some sort of grasp." One observer described the difference in math skills between a McKee Road child, who demonstrated fairly strong performance, and those of the children at Berryhill and the other "low pop schools" as "incredible." In general, the children at McKee had little problem in sequences and functions, the first two tasks in the logical-mathematical section.

The Observers' Role during the Logical-Mathematical Tasks

Mrs. Fox sought to engage all the children in the directions that are administered in the beginning. She called on a variety of youngsters and did so in warm and friendly ways: calling on "the boy in the colorful shirt" or, when she remembered, actually using individual children's names. If a child offered an answer that was awry, she was adept at eliciting the correct answer in a kind manner from the other students: "Who else wants to try?"

The observers at the individual tables also played a very supportive role for the children. In a number of cases, the observers gently prompted students with directions: "Same shape, same number" and highlighted the necessary problem-solving approach:

> Observer: "You figure out how many your minusing" [her hand is touching the child's sequence problem.] "Are these getting bigger or smaller?"
>
> Child: Smaller.

Observer: Right. So you have to figure out how many you're minusing.

Although most children at McKee Road did not seek help, observers still sought to support the youngsters' performances. For example, one observer noticed a child had not completed both halves of one of the number logic problems correctly. She drew it to the student's attention: "Do you see your numbers are different here?" The child said she saw that but "couldn't get the second half.'" For another child who was "stumped," but asked for no help, the observer offered, "Remember, same number, same shape. Is that the same number?"

While the students work on the math problems, the observer at each table records behaviors for each of the students in the students' yellow cards. Behaviors she notes for the various math tasks are contained on a single page, which includes 15 behaviors that can be checked, spaces to jot down the number of correct answers for the various problem types, and two lines for comments. (See Appendix G.)

The Map Task

After completing the math task, the children get a brief break and the observers rotate to a new table. Then, they begin the first task intended to assess spatial intelligence. The "map" consists of a single sheet in the student booklet. The sheet depicts a number of streets that enclose rectangular blocks. Some of the blocks are labelled with locations, such as the word "Playground" and an image of a see-saw and "Fast Food Restaurant" with an image of fries and a soda. In some streets there are problems, or "disasters," blocking the way, e.g, a car crash represented by a drawing of two cars bumping fenders, and a watermain break, represented by a broken pipe and accompanying geyser.

Using a larger version of the map that appears in the student booklet, Mrs. Fox asked children about street names, locations, and problems to elicit some familiarity with the task. She then read them this story:

> One bright, sunny Saturday morning your Mom asks you to help her get some things done. She asks you to buy a loaf of bread and an umbrella. She says you may have lunch at your favorite fast-food restaurant As soon as you are finished, you may go play in the park, but you must return home by 4:00. Before you leave home you hear on the radio that there are some problems on the roads to some of the stores. The problems are: a car accident; a broken water pipe that has flooded a street corner; a fire. Plan what you'll do and then draw the path you will take to get these things done in the quickest way possible so you'll have more time at the park and still return home by 4:00 (Udall, Reid, & Romanoff, 1995, p. 10).

Approximately 10 minutes is allotted for this task.

The Students' Experience During the Map Task

In both schools, students attended to the lead observer's directions. They were eager to volunteer street names, and especially the location of the problems when Mrs. Fox reviewed the map with them. They engaged the actual task as well, tracing a route in pencil around the map in their student booklet. However, an interesting difference emerged in the course of the debriefings. The students at McKee Road actually did what they were supposed to do: they drew an efficient route to the specified places. At Berryhill, many students -- perhaps the majority -- behaved much more like children. Few were interested in getting their mother an umbrella. They preferred to "visit" interesting sites: the fire, the car crash, the flood, the playground. In recounting their route to the observer, they took pleasure in their visits to the disaster areas. They did not appreciate, as did the McKee Road youngsters, that they were supposed to bracket their curiosity and enter into the map/assessment world instead. One of the observers commented that she had never seen so many children stray off course. The assessment team wondered if the seeming uneventfulness of the neighborhood surrounding Berryhill made such disasters more compelling to this group of youngsters.

The Observers' Role During the Map Task

During this task the observer works individually with each child, watching as each youngster retraces his or her route with a crayon over the original pencil route they drew while Mrs. Fox told the story. The individual observers record the order in which the student visited each of the designated places. To keep the children at the table engaged before and after their retracing, the observer encourages the other students to draw a real or imaginary map of a community.

The observers support the students' performance by easygoing questioning. For example, an observer asked a student who stopped before reaching home, "Did you go anywhere else?" "Did you do anything else?"

After working with each child, the observer completes the map section of the child's yellow card. It lists 11 behaviors, such as "Use of road names," "Avoids disasters," "reaches final destination." There is a space for comments and room to note whether a child needed prompting. (See Appendix G.)

Linguistic Tasks

After the map task, Mrs. Fox told the students that they are going to be "working with words" until lunch time. The linguistic section, like the logical-mathematical one, includes several different kinds of tasks:

1. "Contextual clues" consists of eight brief sentences. Each sentence contains a three-letter nonsense word printed in solid caps, and four choices for its possible meaning

are printed beneath the sentence in bold. Children are to circle one of these choices for each sentence. The written directions to the students say, "The sentences below do not make sense. Circle the right word that most likely completes the sentence. Read carefully."

Mrs. Fox helped introduce the task using the examples from the student booklet, the first of these being: "The YAP purred when I rubbed her neck. YAP most likely means: cat fish dog bird." After going through the samples with the students, she read each of the actual questions, along with the answer choices two times, allowing roughly a half minute in between for the children to make their selections.

2. "Categories" presents six problems. Each contains a column of four words that can be grouped into a category. Above the column is a line for children to write in the category title. Below the column is another line in which children are asked to write a word that fits with the given category. The first such question is:

1.________________
peach
orange
apple
banana
____

After all six problems, the children are given a column of six blank lines to make up an openended "category" of their own.

The written directions to the students say, "Think about how these words are alike. On the top line tell how these words go together. On the line at the bottom of the list, write another word that could be added."

Fox introduced the task to the children by telling them, that "we will be looking at groups of words. These words will go together in some way." She then explored two examples from the student booklet with the class. For the first, she asked the class for examples of words that could be added to the list headed by the title "Things that Fly" ("Superman," responded one boy). For the second, she guided the children to provide not only another example, but a title as well. After reviewing the task's requirements, the children had about ten minutes to complete these problems.

The Students' Experience of Context and Categories

As with the earlier math activities, the context and categories tasks are rather test-like. Students are told to cover their work. They work without talking to their neighbors, though they do interact somewhat with the observer.

As with the logical-mathematical tasks, the children at Berryhill seemed more relaxed than those at McKee Road. Perhaps because it had been a long morning, some felt comfortable or tired enough to rest their heads in the crook of their arms, which in turn were resting on their desks as they completed the linguistic tasks.

The Observers' Role During Context and Categories

The lead observer again plays a key role in ensuring the children understand each of the tasks. For example, when a child offered "bird" in response to the sample question "The YAP purred..." Fox lightly questioned "Would a bird purr?" The right answer soon surfaced from the class and directions for indicating it were reinforced by Fox's instructions: "Take your pencil and circle cat."

The observers at each of the tables gently remind children to cover their work or do their own work. In addition, they fill in a half page of the yellow card to note how a pattern is something that repeats over and over. After this, she led the students through two sample sequence problems that appear in the student booklet. Fox then called on children with raised hands, wrote their answers on an overhead which contained the sample problems, and discussed the answers with the class. After this, she drew their attention to the four problems that they must complete and to the fifth openended problem. For this and all the other tasks in the logical-mathematical section, Fox asked if the children had more questions and then told them "You may begin."

"Part II: Functions" also gives four problems for students to solve and then asks them to make up a fifth openended problem of the same type. These problems are comprised of two columns of numbers, labelled "In" and "Out," e.g.,

In	Out	In	Out
2	5	10	8
6	9	9	7
4	7	8	___
3	___	7	5
		6	4

The written directions in the student booklet read: "Find the pattern for each function table and write in the missing number."

To introduce the task to the children, Mrs. Fox used the two sample problems above, which appear in the Student Booklet and which she duplicated using an overhead projector. She explained to the youngsters that "something happens" to each of the numbers in the "In Column" to turn it into a different number in the "OUT Column." When I observed the PSA, Mrs. Fox asked the children to imagine there was a machine that took in one number and did something to make the number come out a different way. What was the thing that the machine did to change the number? Again, answers the many correct answers the children received for the context and category problems, and whether children needed assistance. (See Appendix G.)

Tangrams

After lunch, the students take the tangram task. With some minor variations, this task is nearly identical to the one designed by the DISCOVER team. The PSA employs the same materials: sets of 21 plastic tangram pieces and the tangram booklet developed by DISCOVER. The printed directions to the youngsters are also virtually the same. For example, children are first told that they "each have a bag of colored shapes on the table in front of you. These shapes are called tangrams" (see Chapter 2). As with DISCOVER, the children count their tangrams to ensure they have a complete set. They are shown how to make a square out of two triangles, a larger triangle out of two smaller ones, and a parallelogram by combining the shorter sides of two triangles. This demonstration is largely done by the observers at each table. After this introduction, the children undertake two tasks:

1. They are told to "make a square with as many pieces as you can." (In Arizona they were instructed instead to make a triangle). They are given about 10 minutes for this.

2. They are given the six-page booklet, each page containing various shapes. The children fill in the shapes using their tangram pieces. The directions are nearly identical to DISCOVER's:

> Each page has shapes that are the same as the Tangrams. Be sure to cover up all the shapes on your page. Each page gets a little harder. On most of the sheets you will have to use more than one Tangram piece to make the shapes. When you are finished with each page, tell the adult at your table. Make sure s/he checks your work before you go on to the next page.

> Please continue working until you have finished as many pages as you can. Please make a workspace for yourself so you will not be bothered by others (Udall, Reid, & Romanoff, 1995, p. 16).

Students are allotted 35 minutes, instead of the 30 given to fifth graders in Arizona. Unlike DISCOVER, the PSA does not offer a challenge page.

<u>The Students' Experience of the Tangram Task</u>

Interviews with observers, developers, and teachers of the gifted reveal that students' experience of the tangram task in Charlotte is quite similar to students in Chinle. Children tend to work on the tangrams in a focused way; the task calls for concentration. As in Chinle, some children get frustrated. In fact, at McKee Road, one child got so frustrated she reportedly became "distraught." However, most tend to enjoy both tangrams and the Pablo®. They are reported to say, "This is kind of fun," and they ask the observers when they will come back so that they can do this activity again. Some children also encourage each other, saying "don't give up!" or, "You're never going to get it done if you don't just keep working." The key difference between Chinle and Charlotte is that in Charlotte, very few children finish all of the tangram booklet pages. From the debriefing it appears that only one child in each school completed page six. This may be because the children are younger or perhaps because they work less with visual problem solving than their Navajo peers.

<u>The Observers' Role in the Tangram Task</u>

During tangrams, the observer carries out multiple roles as did their counterparts in Arizona. The observers record the time children take to finish each page, and the order in which children finish the pages relative to their tablemates. Observers also document children's behaviors during the task on the yellow card, which lists 24 boxes of process and product characteristics, including, "works independently," "competes with others," "uses logical strategy for adding or substituting pieces," and "works on several constructions at one time." (See Appendix G.) Observers also monitor the children's frustration level, and they can give out six clues to help the children over hurdles. These include, "You have enough pieces to do this page" and "you may need to take the pieces off and start over again." The clues that are used are also recorded. In Charlotte, unlike Arizona, observers also remind the children to work on their own and note when the children do not.

Pablo®

The Pablo® task used in CMS is similar to DISCOVER's. The materials are the same Pablo® set, described in Chapter 2. As in DISCOVER, the children taking the PSA are asked to make a variety of constructions and to return the Pablo® pieces to the center of the table in between tasks.

Task 1: The observer tells the youngsters, "You may take just a few minutes to make something with the pieces in front of you." They are given about five minutes for exploratory free play.

Task 2: The observer holds up a parallelogram and a triangle. She tells the children, "I am holding two shapes. Use one or more Pablo® pieces to make these shapes." About two minutes is allotted for this.

Task 3: The observer holds up a picture of an animal. She says, "I am holding a picture of an animal. Find one or more pieces that look like an animal. Make your animal on the table in front of you." Approximately five minutes is given for this.

Task 4: The observer holds up pictures of different buildings. She tells the children, "Find one or more pieces that look like buildings. Make the buildings on the table in front of you." About five minutes is given for this.

Task 5: Children are given equal numbers of connectors. They are told to "Make something that moves with as many pieces as you need. Make anything that moves. You can tell us about it if you wish." The children have about ten minutes to do this.

Task 6: For the last task, the children are told "Now you may make anything you would like to make using as many pieces as you want to use." They are given about ten minutes to do so.

The Students' Experience of the Pablo® Tasks

Interviewees reported that almost all the children thoroughly enjoy the Pablo® task, just as the children in Chinle did. During the debriefing, only one child was reported to find Pablo® "too hard." A number of children engaged in pretend play during this activity, making robots, people who go in and out of buildings, and, at McKee Road, an "alien from the planet Exon ... shipwrecked on Jupiter."

The children's actual performances ranged widely. Some made buildings with just two or three pieces. Another, at Berryhill, used 21 and stacked them up vertically. Some made representational work, others' constructions were more abstract or conceptual, for example a "collage" and a "a waker upper." At both schools some children used 3-D and got their constructions actually to move; others didn't. The major difference between the two schools in Charlotte, detected through the debriefing sessions, is that more children seemed to imitate classmates' work at Berryhill.

The Observers' Role During the Pablo® Tasks

In the Pablo® Task, observers are busy drawing each child's constructions on the Pablo® page of the yellow card. They also have 24 boxes they can check to record behaviors for each of the students, among these "received help," "seems excited, absorbed in task," and "works easily and quickly throughout." Other boxes help to record product characteristics, such as whether constructions were three-dimensional, realistic, or had moveable parts. (See Appendix G.) Observers also interact with the children, listening to what the children have to say about their products, asking the children about what their constructions are, and recording what is learned in these exchanges. For example, one observer reported, that a girl had unrecognizable constructions. I'd say, 'What is this?" She'd say, "Craziest monkey on the earth." Information uncovered in such exchanges often materialized in the debriefing sessions and shed light on the students' efforts.

EVALUATION/SCORING OF THE PSA TASKS

The PSA tasks are scored by the team of observers on the same day as the tasks are administered. Following the Pablo® task, the team finds an unused space in the school to confer with each other. For some 15 minutes before the actual debriefing begins, they organize material into each student's Second Grade Classroom Portfolio. (See Appendix F.) These materials include the student answer booklets, their storywriting work, samples of work from the preassessment lessons or the classroom, and their MAT scores.

After the materials are organized into the Second Grade Classroom Portfolios, the team begins working through the yellow cards in alphabetical order. They evaluate all the tasks for each child, before moving on to the next child's work. The usually first evaluate activities in the logical-mathematical realm (sequences, functions, number logic, fluency, story math, and also tangrams), then work in the linguistic area (context, categories, storytelling, storywriting), and finally performances relying on spatial ability (Pablo®, tangrams, the map). Information from the portfolio is called upon as needed.

After each cluster of tasks (i.e., logical-mathematical, linguistic, spatial), is discussed, the child's performance in that area is rated on the 4-point scale: Always evident, strongly evident, evident, or not evident. If a child receives scores of strongly or always evident in two out of the three areas, he or she is officially identified for services by the Program for the Gifted. (The bases for assigning scores is discussed under Condition 4: Clear Scoring Procedures.)

As in Arizona, the Charlotte assessment team spends a great deal of time in debriefing sessions. The two classrooms I observed had 22 and 24 children respectively, and the debriefings lasted between 3.5 and 5 hours. I was told that debriefing sessions typically last close to 5 hours.

In the debriefings I observed, much of the discussion was expedited by considering the number of correct and incorrect answers each child received in the logical-mathematical tasks, and in the context and category sections of the linguistic tasks. Stories from the storytelling task were not always read, but instead summarized. For example, one child's story was "about a rhinocerus." It was "three sentences." In the spatial area, discussion was expedited by considering whether the map task was performed in the correct sequence and the number of pages completed in tangrams.

Perhaps because there are so many tasks for the Charlotte debriefing team to evaluate, and because the nature of the tasks allows for many right/wrong distinctions, the discussions of actual work by students was usually reserved for cases in which the scoring was more ambiguous or borderline. Partly spurred by my comments on this in meetings with Program for the Gifted staff, interviews conducted after I left reveal that the assessment team now devotes more time to considering students' actual products during the debriefings than it did when I observed them. My comments also spurred greater participation by all members of the observation team, whereas during my observations some team members rarely spoke.

ANALYSIS OF WHETHER INCREASED IDENTIFICATION OF UNDERREPRESENTED YOUNGSTERS CAN REASONABLY BE ASSOCIATED WITH THE PSA

In the following section, I consider the PSA in light of the five general conditions needed to make inferences about a student from any assessment. I have argued that these conditions need to be met to associate changes in outcomes with the assessment. I also analyze the PSA against the three conditions needed to associate outcomes from the assessment with MI. (See Chapter 1.) When conditions are not met, I offer suggestions as to how the assessment might be strengthened.

GENERAL CONDITIONS

Condition 1: Children Understand the Tasks

With two exceptions, children do understand what is being asked of them during the PSA. Unlike DISCOVER, which works with many language minority populations, CMS has relatively few such youngsters. Thus, its developers have opted to provide extensive directions preceding each task to ensure children's understanding.

Observers both followed the written directions and illustrated most of them in concrete and interactive ways. For example, in the beginning of the storytelling task, Steve Houser, the PG teacher at McKee Road, asked the students if they have ever told stories, and from whom they've heard stories. Then he asked them, "What do you think makes a story good when somebody tells a story?" After discussing the children's answers, he reviewed story structure: "There's a beginning, and when a person is telling their story, they begin -- A lot of times they begin with a special beginning." The children offered, "Once upon a time...." He then reviewed "the middle part, is the part where you tell what happened. And if there's a problem, you tell about that.... " He explained that, "A story teller also uses the voice, you know, to show the feeling. Have you ever heard anybody tell a story with a giant, and he sounds [uses a deep voice here:] LIKE THIS. Or a little [in a small high voice] little, bitty person. Changes voice...." Only following this extensive, engaging, and illustrated directions, were the children asked to tell a story of their own.

The directions at the beginning of the number tasks were also detailed and involved students. For the sequences task, Mrs. Fox first drew the attention of the students to the overhead where there was a number series: 1, 3, 5, 7, ____ 11. She then said:

> We're going to do two things: look for a pattern today and to see if, when we use that pattern, we can decide what belongs in the blank.... So there are two parts to what we're going to do today. Find the pattern and decide what goes in the blank. When I go from 1 to 3 are those numbers getting larger or smaller?

The class responded, "larger." Then, Fox asked, "When the numbers are getting larger, are we adding or subtracting? She called on a child who said, "adding."

In her interview with me, Mrs. Fox noted:

> Each of the activities as it's introduced has a good deal of teaching or instruction to it. So, as we're teaching a new kind of problem-solving technique, we want to make sure the children understand what they're doing. We have a sample question. And we even go so far in the math as to make certain that every child does get the first one right. If it means sitting there and giving personalized instruction, rather than the instruction that goes on in the front of the classroom ... you take the time to do that.

Alongside clear directions and teaching, observer interactions with the children support students' understanding of the tasks. This is illustrated by the comments of Becky Workman, another PG teacher.

> ...sometimes I ask them, 'what is it that you're supposed to be looking for?' Sometimes I might ask them to read the linguistic part or read the directions to me. Sometimes I think students, if they hear it, [then] they might comprehend it better or see it in a different way. I might refer them, or ask them to look back at ones they have done correctly and see if they can see any similarities. And when children are very, very frustrated, then I might give them some clues, and on my teacher [yellow] card make sure that I mark [that] ...

The exceptions to students' understanding occur in mental math and the map. The distinctions between these two tasks and the others was noted by Romanoff:

> I really think in the storytelling we do get some really good directions, and they [students] know what the expectations [are]. And in our map ... we are missing some of [that].... Also with the story math. I see some holes there, too.

The introduction to the map task familiarizes the students with the map's streets, locations, and disasters. However, unlike many of the other tasks, it does not highlight or preview what makes for a good performance. There is no review, as in the storytelling task, of the elements of good map use. There is also no sample exercise. Such modifications might help the children who do not see the PSA as a high stakes assessment to remain on course.

With regard to story math, Bob Algozzine, a professor at the University of North Carolina at Charlotte who is helping to analyze PSA data, noted that a number of the youngsters just "sat there." It was a "'No lights on, nobody home' number. So, the question is, could they do the mental math? I don't know, because they didn't get [understand] the instructions."

To improve this, the story math directions could discuss the concept of tallying or illustrate different ways people keep counts. It could also provide preliminary exercises to give children practice in the task, akin to the samples provided in many other sections of the PSA. While some improvements might be made on the map and mental math sections, the first condition is generally met: Overall, children do understand what is it they are supposed to do, thanks to clear directions and support from the observers.

<u>Condition 2: Children are Encouraged to Do Their Best Work</u>

In general, the PSA as currently constructed and administered does encourage youngsters to do their best work, though there are elements that might yet be improved in this area.

On the positive front, as noted above, the directions for nearly all the tasks are extremely clear and comprehensive. In almost all of the paper and pencil tasks, examples are given. In the storywriting task, engaging "story starters" are provided to prompt the youngsters. Most of the tasks are also supported through interactions with, and feedback from, the observers. Observers will keep an eye on the students' work to coax them along. As one observer put it, she "made an effort to go back and see what they've done. And I'll go back and say, 'you need to look at this page' or 'you look at that page,' just to give them the same chance'" to do their best work.

Relatedly, as in DISCOVER, the observers in Charlotte seek to establish rapport with the students. For example, Houser reported that it was important "to have some rapport with the student and make them feel comfortable...." Romanoff, who has been a key trainer of the observers, emphasizes that observers must "really try to have a rapport with the kids."

As in Arizona, Charlotte observers are aware of how their gestures and movements communicate with the children. Houser said, "[B]ody language can close it [students' performance] off.... If you have a sensitive and timid child, they won't tell a story if they're intimidated by the person. They're not going to talk or interact in other parts [of the PSA] that they need to interact." As in Arizona, a number of observers mentioned that children who are shy may struggle more with the PSA than they would with a traditional assessment and more so than their outgoing peers.

To support the children's best work during the PSA, children are assessed in a familiar and/or pleasant setting. Children in Challenge Team schools remain in their own classroom and, therefore, worked alongside their classmates. When children are referred from their classroom, the assessment takes place in a familiar setting, such as the art or music room. At McKee Road, the children were assessed in the art room, a rectangular space perhaps the size of three normal classrooms. It had a high ceiling and large windows through which abundant light shone. Houser said, "The atmosphere was, what I would think, an inviting type atmosphere. It was not a cold place to be...."

The ordering of tasks was also done with concern for its impact on children's performance. The observers and designers saw that storytelling was problematic at the end of the morning: The children, like those I observed in Arizona, "...were just like, 'it's time to go to lunch.' They didn't want to tell stories. They wanted to get out the door." Therefore, the designers placed storytelling early in the day. Pablo® was moved to the end, because the materials readily engage children, even at the close of a long assessment.

Perhaps another way that the children's best work is brought out is by the nature of the tasks. It was widely reported that the vast majority of children enjoy the PSA. They regard it as "interesting," or "fun," and "not threatening," according to Udall. Houser reported that "When the students leave, what I've noticed is that they've had a great time and would like to come back." Romanoff's remarks support this:

> I think most of the kids really enjoy what they're doing.... I love at the end of the day, when the kids say, 'Oh, bring your toys back.' And, 'let's do this again!' 'I really enjoyed that.' That is so, to me -- it's just so heartwarming. I'm not kidding. I just get in the car, and say, 'Oh, they had a good day.'

Others observers and designers noted that it wasn't necessarily fun from start to finish for all the children. One said: "... usually the children enjoy the activities. Sometimes when they get to a difficult activity, they can see them as very frustrating. They see that it is a long day with many activities. I think they enjoy it until they reach that frustration point...." Another staff member noted that, while many children enjoy it, "Other children see it as work that's too hard for them. It depends a little on the part of it" [e.g., the particular task].

Reid and Fox also believed that children's experience may vary between schools. In the schools where all the children are assessed, some are not "on grade-level" and may experience more frustration. However, my own limited observations were not in line with this: At Berryhill, where more children were not on grade level, the youngsters seemed relaxed and, for the most part, to be enjoying themselves. At McKee, some of the children were nervous; one girl started to cry.

Clearly, children are experiencing the PSA, and different parts of it, in a variety of ways. It is likely that for many children, the PSA is not uniformly enjoyable. As the description of the tasks reveal, a good proportion of the PSA is test-like -- not typically the pleasantest part of schooling. In addition, for those children who are aware that there is a good deal riding on the outcome, there may be some anxiety involved.

What remains unclear is how, exactly, different feelings about the activities affect students' ability to do their best work: some may thrive under pressure, while others crumble. Some may be lax under enjoyable circumstances, while others may find the same circumstances conducive to good work. What is somewhat problematic is that there may be systematic differences in children's perception of the test, with more affluent children seeing it as high stakes, and poorer children seeing it just as a day that is different from the norm. If this is what children from each setting need to do their best work, so be it. However, I believe this question still needs to be explored and the answer(s) to it used to modify the assessment environment. (See Steele & Aronson, 1995.)

While not all students enjoy all tasks, and though students' differing perceptions of the PSA need to be explored, given the explicit directions and instruction, the rapport, the task ordering, and physical setting that are employed, it is reasonable to credit the PSA with meeting the condition of supporting children's best work.

Condition 3: Evaluators are Trained to Carry out the Work

As in DISCOVER, the PSA observers have a crucial and demanding role, entailing observing and recording students' products and behaviors, interacting constructively with students, and evaluating students' performances.

In asking observers and designers about the challenges of the work, a comment or two emerged about the demands of documentation:

> the most obvious [challenge] to us right now is documenting adequately when you're working with all those kids. Ideally I think you could have even fewer children than five per adult, or you'd have one person administering and a second person documenting.

However, these comments were infrequent, especially compared to DISCOVER, because the complexity of the observers' task has been streamlined. In Charlotte, observers never have more than five children at their tables. Quite often, there are only four.

Furthermore, the demands made by the various recording procedures are also diminished. Observers record students efforts' on tape recorders, and with paper and pencil but, unlike DISCOVER, they are not also wielding cameras or videocameras. They also have only one instrument (the "yellow card"), rather than two used in DISCOVER during the observation and the behavior checklist which follows. Finally, the number of product and process characteristics that the PSA observers are documenting has been reduced on the yellow card to between 15 and 34 (see Appendix G), instead of 90-plus that DISCOVER observers encounter across their various instruments. Romanoff noted that streamlining the instruments was essential to making the PSA workable: "The first year we used her [Maker's] checklist. But everybody was just overwhelmed. It was too much. It was way too much."

A challenge much more frequently noted than documentation was, according to Fox, "being as open minded as possible." Another staffer similarly commented:

> I think the biggest challenge for the observer is probably around maintaining an open, non-biased view of kids. And treating every instance as if ... every child who comes into that room has an equal opportunity of doing well, and of not being thrown by the things that often throw us with kids, like behavior, acting out, I mean whatever.... There's lots of different things that happen so often that can subtly influence the way people see kids.

Open-mindedness is also challenged by behavioral differences the assessment itself encourages. As a third person remarked, "with this type of assessment, you've got to let the kids be a little bit more free. I think that for them [i.e., the observers], they worry about discipline.... I think it's a challenge for them to really be open minded to it."

Given the need for careful observation, documentation, open-mindedness, and the requirement to "encourage, praise, and accept" all children at the table, the observers' task requires training and skill. In Charlotte, the observers' training is commensurate with the challenges involved. Charlotte is also making considerable strides in building a pool of experienced observers.

In the year I observed, and the year preceding it, all observers in Charlotte participated in a training program. At a minimum, the training entails a full day of actually taking each part of the student assessment and also participating from the "observers' standpoint" by giving instructions for each part. A second day of training entails observing the assessment team in action. This minimum training is for all the PG teachers, since PG teachers participate in the observation and evaluation of the students from the school or schools they serve. In addition, both Romanoff and Fox spoke of providing regular feedback to the observers to help ensure quality observations. As Fox phrased it, "We talk with them in an ongoing way about what we think we're doing well and things that need to be improved. So, we have ongoing training."

In interviews, it was quite clear that the PSA designers and administrators in the Program for the Gifted were themselves reflecting upon how to improve observer training. For example, Romanoff spoke often of the need "to do more and more" training. Udall highlighted the importance of developing a library of "training tapes to have examples of exemplars" of children's performances.

Alongside providing and continuing to improve training, the Program for the Gifted instituted the assessment team during the 1995-1996 school year. Prior to this, a school's PG teachers called upon trained substitutes and other PG teachers in the district to help conduct the school's assessment. This meant that observers would be drawing on a limited range of experience gleaned from only a few schools. To make the observations more consistent, a team of 22 trained individuals, either retired certified teachers or substitutes, were hired to work with PG teachers across the district. In 1996, the team was made smaller. As a result, all current team members have a great deal of experience.

In sum, observer training in Charlotte is systematic. All observers are trained. Further, nearly all the observers get regular practice in using the skills the training imparts. (The exception is the school's PG teacher who participates mostly when the assessment team is visiting his or her school and who has some first-hand knowledge about the children being assessed). In addition, the Program for the Gifted staff is reflecting on ways to enhance observer training. Because of all this, the PSA meets Condition 3: it is supported by evaluators who are trained to carry out the assessment.

Condition 4: Clear Scoring Procedures

As the earlier section on "Evaluation/Scoring of the PSA Tasks" revealed, there is a clear organizational process for scoring the PSA: the student booklet, MAT score, writing samples, preassessments, and other materials are assembled into each student's Second Grade Classroom Portfolio. Students' work in each of the three areas (logical-mathematical intelligence, linguistic intelligence, and spatial intelligence) is scored on a four-point scale, ranging from "always evident" to "not evident." Children who score strongly evident or always evident in two out of the three areas are formally identified.

As in Arizona, the structure in which students' work is scored is clear, while the criteria that are used are less so. The actual evaluation criteria that are used are highlighted for each of the tasks in the discussion below. After this, I consider issues that apply across the tasks.

Evaluation Criteria for the Logical-Mathematical Tasks

The designation of a child's strength in the logical-mathematical area is based upon a review of the sequences, functions, number logic, fluency, and story math tasks. The tangram task is also considered for evidence of logical problem solving.

In interviews, observers indicated that one key component of the evaluation was looking at the child's problem solving behaviors.

As one observer and PG teacher noted:

> In the math section, for example, if we had a formula -- that they [the children] had to do [achieve] a certain type of score -- then we're going right back to almost like the standardized test. The idea has been that we look at the children...

Another observer also mentioned a range of behaviors including problem-solving processes and tackling the harder problems. However, her response also highlights the importance of the number of correct answers:

> You can see how many of the problems they get correct. But you can also add evidence to their ability by showing their problem-solving process, or by seeing that they get ... the most difficult problems, even though they might miss some of the simpler ones because of careless mistakes or whatever. So you sort of look for those sparks as well as how many are correct.

As this observer indicated, the initial slice through the students' work in this area was the number of correct answers. The evaluation of these tasks typically began with a review of the scores a child achieved in each of the different tasks.

At the extremes, some children's performances were accorded "always," "strong," or "not" evident based largely upon the numbers of correct answers. This can be seen in the discussion of a Berryhill student, "Kate:"

> Observer 1: 1 out of 4, none out of 4, 2 out of 4. 4 fluency. No story math. By what I'm seeing [from work in the student booklet], she knew what to do. But her math is so weak.
>
> Observer 2: (who observed the child): Yeah. Both of those little girls were in my first group, had some ability. But they just had no math to work with.
>
> Observer 3: Not evident.

At the other end of the continuum, a few children had nearly all the correct answers. While there was discussion of some of their work and strategies, there was little hesitation from the outset that these youngsters were "Always" or "strongly" evident. Of a McKee student, "Jeff," an observer reported:

> ... In math he was 4 out of 4, open [i.e., he created a problem of the same type, or "openended"]; 4 out of 4, open; 4 out of 4, open. [reading

> observer's comments:] 'Had no problem with this. He worked hard. He got them all right.' He had 21 fluency.... And he had no answers in his story math. He had the right method, but he added wrong.

The observers briefly considered whether to score this work as strongly or always evident, and decided on always evident.

Beyond considering the number of right and wrong answers, one prominent variable that influenced the observers' decision was the presence of an open-ended problem. Open-ended problem solving both reflects Maker's influence (see Chapter 2) and is central to the district's definition of giftedness, described earlier. ("...When presented with an open-ended or challenging problem, extraordinary problem-solvers demonstrate creativity, critical thinking, and task commitment in order to reach a productive solution.")

The absence of an open-ended response influenced scoring for "Amy," a child at McKee Road. Amy's debriefing began, "4, 3, and it looks like 3, with 12 fluency and no story math...." The observer's comment for Amy: " She worked very hard on the functions and seemed to understand the concept. She had help on sequences and number logic. She keeps on working. She never quits." Of this child, Fox said, "It concerns me that we give somebody like this strongly evident, when there's no attempt at open-ended." In contrast, Charles at Berryhill, scored "2 out of 4, 2 out of 4 and an openended, 1 out of 4 with an openended, 14 in fluency, 2 in story math." He ultimately was given strongly evident in the logical-mathematical portion of the assessment, partly because the observers appreciated that "he really did two openended."

Another variable that influenced the observers' scoring was a child's need for help. This is clear in the following discussion of "Brian," a Berryhill student, who scored "3 out of 4, 2 out of 4, 3 out of 4, 22 in fluency, and 1 story math."

> Observer 1: Any help?
>
> Observer 2: No. No. I was so busy helping the other three [at the table], I didn't have time for the poor child. This is his work.

Ultimately, the team accorded him "strongly evident."

Another child at Berryhill, who got some right answers (2 out 4, 1 out of 4, 2 out of 4), was "not evident" in part because she "had a hard time. She needed a lot of help." At McKee, "Amy" was given "evident" instead of strongly evident both because there were "no opendended at all in there" and "the fact that she needed help ... more than once."

Alongside openended problems and the need for help, a frequently mentioned criterion was evidence of problem-solving strategies. The presence of a strategy most often came up in the discussion of the story math and the number fluency task. In story math, the observers looked to see if a child tried to record the objects that were in the story, using either words, pictures, objects, or fencepost talleys. On the number fluency task, it was noted when a child used multiples of 10 to generate his equations: "He was having his strategy. And when he did his fluency, he did, again, the multiple numbers of 10: 70 - 60 = 10."

However, the presence of strategies did not seem to tip the scales for students whose math skills were deemed weak in some way. For example, "Jake," a Berryhill student, used the strategy of multiples of 10. He also used small scraps of paper to represent the objects in the story math problem. But "large numbers posed a problem" for him, and his "numbers [basic number facts] maybe are just not sophisticated enough." He scored 3 out of 4, 2 out of 4, and 2 out of 4 on the first three math tasks, with 14 in fluency and no correct answers in story math. The team acknowledged that Jake's case was a hard one, but they ultimately decided his performance supported a score of evident, rather than strongly evident.

Hallie at Berryhill had the same scores as Jake: 3 out of 4, 2 out of 4, 2 out of 4, 14 in fluency and no story math. She was also given evident, even though she had a strategy in story math including using plus signs and writing down the words for the objects to be counted.

At McKee, the presence or absence of a strategy was mentioned, but again did not seem to influence the observers' basic evaluation. These children tended to be scored high, whether or not they had a strategy. For example, Jeff, the McKee student had all the right answers in the first three math problems, plus openended problems for each. He didn't answer any of the story math problems correctly, but he did have a strategy. He was given always evident for the logical-mathematical tasks. Tammy completed the first two tasks correctly and provided an openended for each. She answered only one number logic problem, had 12 in fluency, and 2 in story math. She showed "no strategy that I can see," yet, after some debate, she was still designated "strongly evident."

Other items and criteria were called upon in the debriefing discussions for the logical-mathematical portion of the PSA. For example, the team looked at how many tangrams the child completed and evidence that the child used logical problems solving processes in tangrams. They also looked at the children's MAT scores, and the teachers' evaluations from the preassessment lessons. In fact, most of the other criteria listed in the logical-mathematical section of the yellow card (See Appendix G) were mentioned at various points. However, like evidence of strategy, these additional criteria rarely superceded the first three mentioned: the number of correct answers, the child's need for help (which may influence the perception of the number of correct answers), and the presence of openended problems.

Evaluation Criteria for the Linguistic Tasks

The evaluation of linguistic intelligence is based on the team's review of performances in the context, categories, storywriting, and storytelling tasks. In interviews, designers' and observers' emphasis was clearly placed on the storytelling and storywriting tasks. In fact, in interviews no one mentioned how the context and categories tasks weighed into the scoring, except to say they were "looked at" as part of the effort "to get a fair idea of that child's linguistic ability."

The review of linguistic tasks typically began with a quick summary of how many correct answers a child supplied for the context task, and the number of titles and items the child supplied in the categories task. After this, there was usually some discussion of the stories the child told and wrote. The discussion for Winnie, a child at Berryhill, was:

> Observer 1: "5 out of 8, 4 out of 6, 4 out of 6 in categories. No open-ended. She had a monkey that grew some long legs. But that's all that she says. She have a teacher story writing?
>
> Observer 2: 2-plus.[10]
>
> Observer 3: Storywise, this is not good. That was not good. This is not her medium.
>
> Observer 1: So-so. [agreeing that her performance is not strong]

Observer 3: Evidence based on that?

Observer 1: I can give maybe evident?

Observer 3: You can?

Observer 1: Maybe. What does everybody else think?

Observer 4: What are we basing it on?

Observer 1: 5 out of 8, 4 out of 6, 4 out of 6, not much of a story, and a 2+ here [on the preassessment writing]

Observer 3: Alright. Not evident.

Observer 1: Is that alright?

Observer 4: Sure. I don't see it [evidence of strength] being backed up with anything.

For Adam, another Berryhill student:

Observer 1: 5 out of 8, 5 out of 6, 6 out of 6.... He did a zebra I guess?

Observer 2: Yeah. I think he had the zebra. Oh, he started off by saying, I know: Oh, he gave the differences between the zebra and the unicorn. And he talked about the horn on the unicorn and he talked about several instances. He gave examples of them playing together. He didn't really have a great story here. He itemized a lot of the things they did in playing. And that was about it.

Observer 1: His story writing?

Observer 2: 1 on story. [the teacher's score of his preassessment storywriting]

Observer 1: 1 on story.

Observer 3: He was enjoying telling it though. [enjoying the task is an item on the yellow card.]

Observer 1: On his categories -- oh, he's the one who pulled out his crayons.

> Observer 3: Yeah. And copied them down. [i.e., he copied the color names from the crayon labels to make a category.]
>
> Observer 4: I wondered [how he got the category]: red, orange, green, turquoise, blue.... He was using his head.
>
> Observer 1: Do you see any evidence? I mean, I don't see him as weak as some of the ones we've seen. But is it enough to give him evidence? [others gesture disagreement]. No? No evidence.

In the two examples above, the storytelling was regarded as not strong. However, notable differences in the categories task -- even the comment that a child was "using his head" to create a response to the openended category problem, didn't influence the outcome.

The difference in scoring that a good story makes is highlighted by the child who told the story about the panthers mentioned earlier in this chapter. At McKee Road, several children had completed all the context and categories task correctly. This girl did not, but was still given strongly evident.

In scoring the storytelling task, observers commented that they were guided by the behaviors listed on the yellow card. One observer and PG teacher said of both storytelling and storywriting:

> You look for something that is out of the ordinary as far as story line. You look for humor. You look for detail. You look for advanced vocabulary. You look to see if the story has a beginning, and a middle, and an end; if it's all pulled together at the end. That's basically it.

According to this observer, these behaviors are both on the checklist and emphasized in training. Another observer also stated that story elements on the checklist guided the evaluation, "including the story has a clear introduction and a conclusion, the story has a clear sense of place, the story includes action." This observer added that no one of these is give more weight, because "there are so many elements in there that come together to make a whole."

In sum, the assessment of the linguistic area is governed by the storytelling component, and to a lesser extent, the writing component. The criteria governing these tasks' evaluation are clear to the observers: based on interviews and debriefing data, they actually do use most of those on the yellow card. (See Appendix G.) What remains unclear is why the context and categories tasks are included, when they appear to have only minimal impact on the evaluation of ability in this area. This issue is considered again in the discussion of intelligence-fair and domain-based conditions, below.

Evaluation Criteria for the Spatial Tasks

The tasks that contribute to the assessment of a child's abilities in the spatial area include Pablo®, tangrams, and the map. Criteria for each of the tasks is discussed below.

Tangrams: The key criterion in this task, as it was with the DISCOVER team, was the number of pages finished. Other considerations included the amount of time it took for the child to do the work, whether the child could complete page 3, and whether or not the child had help. As one observer described it:

> We look at how many pages they have completed, which is not really what we're supposed to do, but it's certainly some measure of how quickly they were able to solve the pages and go through.... [T]here are some pages that are more difficult than others. I think page 3 is very hard, and if the child works a long time with that page and then completes it, I think that also says something else about a child's persistence, and desire, task commitment.... It is very good, I think, that we keep a count of how many minutes it takes for each one [each page], because that is a very comparative and competitive thing. So a child that does really well at one table might not look so good at the next table, unless you know how quickly they completed the pages.

Evidence from the debriefing sessions emphasizes that the observers do focus on the number of pages, speed, and completion of page 3. At Berryhill, the discussions of tangrams ran from a number of very brief comments, such as "she got absolutely stuck on page 4" to somewhat more descriptive summaries:

> On page 3, she got frustrated. Because she looked around and realized that she's toward the end of the group [i.e., among the slowest at the table], and she's not happy. She did work on breaking [the problems] down. I worked through with her. She only worked 2 [i.e., completed two pages].

At McKee, the evaluation also emphasized the number of pages and the speed with which they were completed: "Finished through page 4;" "this child who was on page 5, was first or second [place] in everything she did." The discussions at McKee were generally also quite brief.

Although the observers had some 30 behaviors and product characteristics listed on the yellow card for tangrams that they were trained to look at, in essence they evaluated the task by considering a small number of these: number of pages (and specifically getting through page 3) the speed with which the pages were completed, and help received. This is quite similar to the approach taken by DISCOVER observers in evaluating tangrams.

Pablo®: In interviews, the observers provided several criteria that they considered in evaluating Pablo® constructions:

> [W]e look for something the child has made that is different from the other children, something that they can explain. I look for three-dimensional work, but don't always rule out the two-dimensional work as not creative. You look for the details involved. Sometimes you see the child has put down 30 Pablo® pieces, but they're just like they keep on going -- they didn't know what it was. They just liked the colors, but it does not have a shape or a title.

This comment, and the actual conversation in the debriefing sessions, indicate that the Charlotte observers relied on a limited set of criteria to evaluate the Pablo® tasks. Like the DISCOVER observers, those in Charlotte put great emphasis on the presence of 3-D work. However, unlike their counterparts in Arizona, the Charlotte observers also placed considerable weight on whether something was representational and whether it was unusual or unique. In Charlotte, the observers were also less likely to credit constructions that had been copied or inspired by others.

Of Winnie, a Berryhill student:

Observer 1: Her pablos. She copied another student's clock.

Observer 2: Remember the good clock? The one [by student's name]. That was her cow [points to drawing]. She had the head, you know, perpendicular to the cow. The body held it up. That was a plain building. Just a robot. She might've copied that too....

Observer 1: What kind of comment would you make about her pablos?

Observer 2: Um, not interesting.

In contrast to Winnie, a child whose work was judged strongly evident independently produced unusual, 3-D constructions:

Observer 1: ... His building was kind of good. The other ones -- His other ones --

Observer 2: His laser gun --.

Observer 1: Yeah. You know they're 2 pieces, 3 pieces. Nothing fantastic.

Observer 3: Is the bird with jellyfish -- catching the jellyfish in his mouth?

Observer 4: Wow, that's pretty good.

Observer 2: It's unusual. I mean it looks like one.

Observer 3: What was his building? 21 pieces on his building.

Observer 1: He just put together, going up vertically. And they all fit real nice. And there is some 3-D.

Observer 3: I like his ladybug.

Observer 1: He had a cute ladybug.

The value of representational work is also highlighted in the work of Amanda, a child from McKee Road. Here it seems to count even more heavily than 3-D: Her work in this area, though 3-D, was harder for the observer team to judge, because her 3-D structures did not depict or reflect what she called them.

Observer 1: I had her for Pablos. The first little thing was cute. It was a mouse.... The toystore was the second one. This was kind of a neat building. Because it was very, it was very symmetrical. She paid a lot of attention to the detail. Now, when it got more open-ended, she got weaker. Because the next thing she did she called "a waker upper." It was supposed to wake you up in the morning. But she couldn't tell me why this piece was here. What it was supposed to do. It was just a waker upper. It had a lot of pieces, and it was very 3-D. It went off in all kinds of directions. But with no apparent purpose for going off in all kinds of directions. And you saw her alerter: 'It alerts people to fire or if something bad happens. It's very loud.' And there again, it had lots of pieces going off in different directions, but nothing that you could look at and say: aha!
...
Observer 2: She basically put a bunch of pieces together and gave it a name.

Observer 1: Uh-hm.

Some of the best Pablo® work at McKee was done by "Trevor." This work was very representational, three-dimensional, and showed great attention to detail. In addition, this child led his group, as opposed to copying others.

Observer 1: I mean, everybody followed him. And he was good. This was a deer.... It had antlers and eyes. He fixed the eyes in such a way that they sat up. I mean he was that careful in how he placed [pieces]. And even though he didn't have connectors at that point, he had those eyes

> standing up. And he had the legs precisely placed. And then, let's see: The building wasn't outstanding. But it was very symmetrical. It was very carefully placed. And this was an arch over the door, covering the door from the rain.... And this was the clock. It was some kind of office building. It had a clock.... It had a lot of roofs.... Now when he got to the something movable and the open[ended ... he did a plane. And it was *excellent*. I mean, he had the wings coming out. He had the undercarriage where you had the semicircles [depicting wheels]. So that it sat at an angle when you set it down.... He had the tail thing that goes up and then goes back. It was very good.

In sum, for the Pablo® activity a limited number of product and process characteristics guide the observers' scoring. Strong work is generally regarded as being unusual, but representational, three-dimensional, and independently done.

The Map: The criteria for evaluating the map were clear and straightforward. The child had to go to the four specified areas in the correct order, return home, use the most efficient route, and avoid all disasters. Characteristic discussions of the map at Berryhill were: "Spatially his map: ... Home, grocery store, department store, french fries, playground, car crash, [broken water main] pipe, fire, home." Or:

Observer 1: His map: 2,1,4,3. He wandered.

Observer 2: Yeah. His map made no sense whatsoever.

In contrast, at McKee Road: "Her map is 1, 2, 3, 4, home;" "His map is 1, 2, 3, 4, home. He used road names. He used left and right.... He used an expeditious use of time."

Criteria such as using road names, using left/right directions, and using place names appear on the yellow card. However, as illustrated by the review of the first child at McKee ("...1,2,3, 4, home"), these other characteristics did not influence scoring. The main criteria were completing the course without wandering or visiting the disasters.

Though no one articulated how the evaluation of the three spatial tasks are weighed together to yield a score in that area, the data from the debriefings indicates a clear rule of thumb. First, not surprisingly, if all three tasks shows no evidence, the child is given a score of not evident in the spatial area. This is illustrated by Katie: Her Pablo® constructions were "very simple," she finished only two pages of tangrams without help, and her map was out of order.

If a performance with either Pablo® or tangrams is solid, the child is given a score of "evident." This is illustrated by Berryhill's Winnie, who received "evident" with Pablo® work that was "not interesting," a disordered map, but four completed tangram pages. Similarly, Gerry's Pablo® work was mostly "just stuff," his map was disordered, but he finished 3 pages on tangrams.

If two or all three tasks appear strong, the child is usually scored strongly evident. Thus, Hallie at Berryhill was labelled strongly evident based on a "fine" map, only "fair" Pablo® work, and "very strong" tangrams. This was also illustrated by Trevor, whose map was correctly ordered but used an inefficient route, whose Pablo® constructions (including the airplane described earlier) were very strong, and who quickly completed four pages on tangrams.

The criteria in actual use for the spatial tasks are clear to the observers. They are able to use only a few of those listed on the yellow card to examine and evaluate students' performances in Pablo®, tangrams, and the map. They also appear to use the rule of thumb just described to combine performances across the three spatial tasks to provide the overall scoring in this category.

Issues pertaining to scoring procedures across tasks

Several issues are common to the scoring of all the tasks:

"All clear:" The criteria that are actually used are clear to the observers across the tasks. No discussions or interviews revealed questions about the meaning or relevance of any of the product or process characteristics. This represents a definite advance over DISCOVER's checklists.

"Many called, few chosen:" Far more information is collected than is used to evaluate this work. For example, on the yellow card, though no one questioned the value or meaning of any of the items, except for the storytelling and storywriting tasks, Charlotte observers still relied on just a few of the 15-34 characteristics that are listed. Leaving aside storytelling and storywriting, observers used between approximately 12 and 27 percent of the characteristics listed on the yellow card. Here, as in Arizona, the observers may be "satisficing" (see Chapter 2): accomplishing a rather complex task by drawing on a limited amount of information.

At the time of my visit, efforts were underway to pare away additional behaviors and characteristics, in part to reduce the load on the observer team, and in part to make the card reflect those characteristics that are actually at play in the evaluation. However, the yellow card also provides a protective mechanism: it serves as evidence to back up decisions. Several times, observers had to make sure that the card itself justified, or could support, their decision. For instance, after scoring a child as strongly evident in math, despite the child's careless calculation errors, the observer inspected the yellow card to make sure she "certainly documented" why the student was given this score. In another case, the observers were not much impressed by a child who told a story about a jaguar king who forces other animals to cut up fruit for him. Though the child had many checkmarks, her story line or rationale was weak. The observers, then noted on the yellow card that the story did not follow the prompt and that it didn't follow a logical sequence.

Thus, paring away the checklist to the smallest possible subset may eliminate some of the evidence that shores up the decisionmaking process. It may take away some of the subtleties noted by an observer but not discussed in the debriefing, and it may undermine the security observers feel in their decisionmaking.

Another source of surplus information is the data in the student's Second Grade Classroom Portfolio. Information from the preassessment lessons, the MAT, and the teacher recommendation was occasionally called upon, but generally did not influence the evaluation of linguistic, logical-mathematical, or spatial ability. Typically, this additional information was accepted when it supported observers' impressions of the work and rejected when it contradicted the observers' evaluations. For example, a Berryhill student whose Pablo® work was "really unique," but who also had a disordered map and only two pages of tangrams, was labelled evident in spatial intelligence. This decision held, even though she had the highest possible score on the MAT, a test of figural problem solving. Another Berryhill student was given evidence rather than strongly evident in linguistic intelligence, even though she had a very strong storywriting assessment from her classroom teacher. This wasn't given much weight, because it wasn't clear if the teacher "leaves it there for a morning's work or for three days." Again, I believe this represents the observer team's need to get through a complex and time bound task in the most efficient way possible, i.e., by satisficing. However, it is possible that it also reflects a bias toward the tasks that they have observed and administered themselves.

While much of the information from the students' second grade portfolio is not drawn upon, the additional information represents a developmental process on the part of the PSA's designers. It was spurred in part by a desire to gain a more complete picture than the DISCOVER assessment permits (hence the additional tasks). It was also spurred by discomfort with 'one-shot' evaluations that occur in DISCOVER and psychometric tests (hence the preassessment material). The designers are now in the process of training the observers to review all the material more thoroughly and draw upon it in their evaluation.

"It floats:" In Arizona, observers knew that the reference group for scoring was supposed to be the classroom, even though they had trouble following that guideline. In Charlotte, the Ivory Soap adage describes the reference group, because observers and designers had a hard time explaining who or what set the standard for a particular score.

> MK: How do you set standards, criteria --?
>
> Observer: It's very hard. We've wrestled with that the whole year. We really can't say that the level of instruction is so low in the school that a child who performs relatively better than the others -- that you don't belong in the program, or vice versa. It sometimes -- I sometimes wake up with that issue in my head. It's been a bit of a battle to say we're going to look at a single standard, but that these schools vary tremendously.... At the beginning of the year, we were leaning in the low-pop schools toward the classroom standards. But we were caught up short [by a dispute among PG teachers]: 'Are you judging them from we're they're coming from or [by] the same standard?' There has to be the same standard -- or do we look at where they're coming from? No one is giving us an answer.... The logical side of me wants to have an answer: 'This is what you need to do.' But no one wants to do that, to say, you know, 'don't do that' or 'do do that' when you go from school to school.

The floating nature of the reference group or standard comes through in the variety of different standards used. Sometimes, as with tangrams, task-specific standards weighed heavily. This is clear in the importance acorded to completing page 3. At other times, the reference group is the students' classroom. For instance, when the assessment team began the debriefing at McKee, the observers first tried to figure out for each cluster of tasks "what would be a minimum requirement for strongly ... in this group." The classroom standard also seemed to be at play at Berryhill, when one student's work was compared to another, e.g., "she definitely did more than the rest of them." However, at still other times, the reference group was the school or the district -- specifically, the capacity to perform in the gifted program in any school in the district. One person voiced both a district-wide standard but also the recognition that students' classrooms and schools impinge on this approach.

> We struggle with what's the context of our school system. Because these children move from place to place, and they need to participate programmatically, no matter where they come from. And on the other hand, you've got the context of their school and their classroom.... But if a child from Berryhill moves to McKee Road, they need to be able to hold their own with those kids.

This person recognized that this issue is "one of those weak spots of our assessment."

Though underdefined conceptually, the reference group (or groups) nevertheless appears to be working from a pragmatic viewpoint: Several sources reported that the group of children identified in the previous year for the participation at the third grade gifted magnets "was the strongest group of kids we've had." Somehow, by keeping in mind both a district-wide standard, while making some mental adjustments for school context, the assessors have been able to select a much more diverse group of students who still perform well in gifted education contexts. As Carol Reid put it, "The only proof is in the pudding." It is also possible that over the years since the PSA began, PG teachers have become better able to teach to the strengths of the children selected on the new assessment.

To summarize, this lengthy exposition reveals that the Charlotte group has made strides in simplifying the long lists of process and product characteristics. Furthermore, observers reveal a general consensus about which of these characteristics are actually applied in the evaluation for all the tasks. In these ways, the bases for scoring the PSA have advanced over the pioneering efforts of DISCOVER. Yet, because the reference group or other standard could not be articulated, it is not clear how these characteristics yield one or another score.

Since the time of my observation in October 1995, the PSA designers have begun developing rubrics against which the tasks can be measured and scoring categories assigned. When sensible rubrics are in place, then the question "compared to what?" should be evident to observers both inside and outside of Charlotte. Given that the reference group or standards may still float, it is not reasonable to credit the PSA with meeting the condition of clear scoring procedures. However, it does seem that Charlotte's assessment is on its way to anchoring its scoring through the use of rubrics and ultimately to meeting this condition.

Condition 5: Observer Reliability

Formal efforts to evaluate observer reliability for the PSA have yet to be undertaken. At this point, there are only informal indicators that some reliability exists or is being developed. One source of evidence is that when a child's PSA is disputed either by a parent or the PG teacher, it is sent for independent, blind review to five or six other observers. In the roughly half-dozen of these instances, the blind reviewers have never overturned the initial team's evaluation.

A second indication that observer reliability is being developed comes from the assessment teams. Prior to the teams, one observer felt that the assessment itself varied from place to place. Children's work was "being interpreted differently in the different schools. Because in every school it was a different team." The observer pool was developed in order to "add some stability to the testing" and, per another designer, because administrators in the Program for the Gifted wanted the PSA "to be more consistent."

An additional source of evidence comes from a key correlate of reliability that Griffiths' (n.d.) found in her investigation of DISCOVER, namely observer experience. (See Chapter 2.) Giffiths found low inter-observer reliability between novice observers (those with less than 15 observations) and the highest reliabilities between observers with at least 30 observations. The creation of the PSA observer pool in Charlotte has yielded a group of highly experienced assessors. In 1995-1996, there were 22 observers, most of whom worked on observer teams several times a month. By the end of the 1995-1996 school year, 14 members of the observer team logged 30 or more observations. Another eight had between 15 and 30. There are no longer any novice-level observers in the group. By the end of 1996, each of the nine observers used in the 1996-1997 conducted more than 75 observations.

Because no formal studies of reliability have been undertaken, it is not possible to say that the PSA meets this condition. At this point, there are only preliminary indications, from independent, blind reviews of students' folders, from the creation of the observer team, and from the level of observer experience, that the PSA is on its way to achieving observer reliability.

MI-SPECIFIC CONDITIONS

The five conditions discussed above are needed to to make inferences about students' abilities from any assessment. They are therefore needed to link changes in who is identified with the assessment (See Chapter 1). In contrast, those below are needed to associate the assessment with MI.

Condition 6: Assesses Abilities Beyond the Boundaries of Traditional Tests

Even though the developers assert that their assessment draws on MI theory (e.g., Reid, Udall, Romanoff, & Algozzine, in press), the tasks used to identify children do not draw on most of the intelligences encompassed by the theory. Like the DISCOVER assessment, the PSA assesses the same three abilities traditionally measured by psychometric tests: logical-mathematical, linguistic, and spatial.

That the developers of the PSA still feel their work is linked to MI may be attributed to several explanations. First, their work was *inspired* by Gardner. As noted in the opening of this chapter, when Superintendent Murphy convened the task force on gifted education, its members were attracted to Gardner's ideas, and they sought to use an assessment modeled on the theory. Second, the model they first adopted was Maker's, and, as indicated in Chapter 2, Maker claims to be applying MI in her assessment. Third, MI does appear to be used within the county. There is professional development around MI for gifted educators and regular classroom teachers; one school in the district has an MI focus, and teachers at one of the high schools were embarking on an MI curriculum while I was visiting. Thus, MI is present in Charlotte, though little manifested in the PSA itself. Finally, because the PSA incorporates some of the hands-on activities that DISCOVER employs, it is viewed as more intelligence-fair than traditional tests. (This assessment's intelligence-fair nature is examined in the next section.)

Why isn't MI being used? This issue is explored fully in the last chapter. However, in brief, one reason is that the curriculum in the gifted program does not yet encompass opportunities for the range of intelligences to be drawn upon. As one designer and district administrator stated, if "we identify an interpersonal child, what are we doing in class for that child?" Commentators on assessment often note that testing drives the curriculum (Frederickson & Collins, 1989; Haney, 1989; Wiggins 1989, 1993a, 1993b; Wolf, LeMahieu, & Eresh, 1992). This appears to be a rare case of curriculum driving assessment!

Condition 7: Intelligence-Fair

The PSA does have some intelligence-fair tasks: the Pablo®, tangrams, and storytelling activities. These allow children to demonstrate their problem solving ability in media other than paper and pencil. Yet, as a whole, the PSA is dominated by paper-and-pencil, or "second-order" tasks. Second-order tasks draw not only on an ability itself, but on the capacity to represent that ability in notational form. (See Chapter 2.)

In the linguistic area, both first- and second-order performances (storytelling and storywriting, respectively) contribute to a designation of strongly evident. As noted above in the discussion of Condition 4, context and categories, two paper-and-pencil tasks, have little bearing on the evaluation of students in this area. When I observed, storytelling carried more weight than storywriting. However, follow-up interviews in the fall of 1996 indicate that storywriting is taking on increasing importance in the 1995-1996

PSA. Thus, it now appears that children need to perform well in both first- and second-order linguistic tasks to be strongly evident in the linguistic area.

In the logical-mathematical area, sequences, functions, number logic, and story math are all second-order tasks. Only tangrams, to the extent that it is considered alongside these, is a first-order task that draws on logical-mathematical intelligence. Therefore, to qualify as strongly evident in the logical-mathematical realm, a child must be able to translate logical-mathematical strengths into notational form. In addition to the notational demands in the logical-mathematical area, there is a heavy "verbal load" in the story math task. To succeed on that task, children must be able to follow three rather long problems conveyed largely through spoken language.

The spatial component of the PSA includes two intelligence-fair tasks: tangrams and Pablo®, and only one second-order task: the map. It is the only area of the assessment in which a child is likely to be identified as strongly evident on the basis of intelligence-fair tasks. Because children must be judged strongly evident in at least one of the other two areas to be identified as gifted, it is not possible to say that the PSA is an intelligence-fair measure.

Given the weight placed on notations and language, observers in Charlotte are much less likely than their DISCOVER counterparts to give students the "benefit of the doubt." In Charlotte, the burden of proof lies much more with the student. As one observer commented:

> I think one challenge is in seeing a child having a spark: seeing them show in some way that they do understand, and are able to think on a higher level, and yet not seeing that consistently through an intelligence [set of intelligence-related tasks]. And it makes you want to say the child

> has potential, but you can't always document ... the insight or intuition that you might have about the child's ability....

For an observer to document strengths in two of the three areas needed for identification, a child needs to demonstrate his or her ability in notational form.

Why has Charlotte veered toward more traditional paper-and-pencil measures? When they began developing the PSA, the designers tried "to get as far away as we felt comfortable from paper and pencil. We wanted hands-on activities." Yet, a strictly hands-on test proved problematic, because it ignored the demands of the gifted program curriculum. Therefore, while keeping hands-on tasks, "We have also acquiesced to the [PG] teachers' comments that in order to perform in the program, in the classes and to do the kinds of intensive research and work that is anticipated for them, they need to have some of those [notational] skills." Thus, while intelligence-fair practices are understood and valued, they have been submerged under pragmatic pressures to select students who can meet the PG teachers' expectations.

The emphasis on notational skills may undermine the Program for the Gifted's goal of increasing minority representation. There is ample evidence that children from less affluent and educated households acquire literacy skills later than those from more privileged backgrounds. For example, the differences in mothers' education, like those noted for Berryhill and McKee, correlate with early differences in literacy skills (Daiute, 1993; Snow, 1991). Especially at the beginning of second grade, when the assessments I observed were held, poor and minority children were almost inevitably going to be infrequently identified on such a notationally laden assessment.

To their credit, in 1996-1997 the PSA designers shifted the assessment schedule, so that children at the "low pop" schools now take the PSA after those who attend schools with more privileged populations. This leaves several additional months for the PG teachers to provide preassessment lessons and collect additional information about the students. The new schedule also provides these youngsters with further immersion in literacy-rich environments, which should prove beneficial on the PSA. However, the PSA itself might be revised to include more intelligence-fair tasks, so that students, especially at this young age, need not meet the burden of proof notationally in two out of the three areas.

Condition 8: Domain-Based

In Gardner's theory of multiple intelligences, intelligence becomes evident only in culturally valued practices or "domains" (Gardner, 1983, 1991a; see Chapter 1). Some of the PSA tasks do meet this condition, including the storywriting, storytelling, and map tasks. Others, such as tangrams, Pablo®, context, and categories do not reflect domain-related practices. In fact, one designer reported that the latter two tasks were imported from some of Robert Sternberg's assessments. As noted in Chapter 1, Sternberg especially values the role of novelty in assessment (Sternberg, 1985, 1988). As noted in Chapter 2, such novel tasks are fundamental to psychometric assessment. Thus, while part of the domain-free nature of the tasks arises from historical links to DISCOVER (i.e., tangrams and Pablo®), other domain-free tasks have been consciously selected.

Using domain-free or novel tasks enables the Charlotte team, like its DISCOVER counterpart, to maintain ties with the psychometric mainstream. However, as with

DISCOVER, the presence of such tasks makes it harder to argue that the PSA is a domain-based assessment and diminishes its tie to MI theory.

Such a link is not only a theoretical nicety. One of the key benefits of using domain-embedded tasks is that they provide a way to anchor evaluation in meaningful criteria. For example, Spectrum assessments (see Chapter 1) used domain-based tasks and materials in order to ascertain children's strengths. By assessing children's spatial ability partly through the youngsters' visual art work, they could apply art-based standards about line, composition, and expressivity to the evaluation of the work (Krechevsky, 1994). Using domain-based tasks may help the PSA developers in the effort to evolve meaningful rubrics.

CONCLUSION

Charlotte's PSA is an assessment that was greatly influenced by the work of DISCOVER. Yet, the PSA has evolved substantially from that starting point. This evolution enables the PSA not only to meet the first two conditions (children understand the tasks; children are encouraged to do their best work), but also the third condition: that evaluators are trained to carry out the work.

In addition, the designers of the PSA have taken pains to eliminate excessive product and process characteristics from their observer instrument, the yellow card. The characteristics are now clear and reasonably concise. The designers are currently developing rubrics that highlight characteristics associated with different levels of performance on their four-point scale. When these rubrics are in place, this should enable all concerned to answer the question, "Compared to what?" The PSA, therefore, has not yet met the fourth condition (clear scoring procedures), but is on its way to meeting it. The same situation holds true for observer reliability, the last of the five general conditions. No formal studies of observer reliability have been made. There is some informal evidence supporting observer reliability from independent blind reviews of children's work. In addition, observers have extensive experience, which correlates with observer reliability, at least among the DISCOVER team (Griffiths, n.d.). To ensure its observers maintain their skills, the designers have gone so far as to cancel the contracts of observers who did not participate frequently enough.

At the same time that the PSA has gotten closer to achieving all the general conditions, it has veered farther from its theoretical underpinnings in MI. It does not assess abilities beyond those traditionally tested (Condition 6). While it has some intelligence-fair tasks, overall the PSA does not enable children to be identified on the basis of their performance on such tasks (Condition 7). Finally, it is largely not a domain-based assessment (Condition 8), and it has become less so over time.

The fact that the PSA does not meet any of the MI-specific conditions does have some costs. Clearly the rhetoric and the reality of the assessment are out of alignment. A more significant problem is that the assessment may not be detecting as many underserved youngsters as possible. This is highlighted in the above discussion of intelligence-fair measures. It is also true that by looking at the narrow range of abilities that currently mesh with the gifted curriculum, youngsters with gifts in other areas go unrecognized. It may be the case that Charlotte could identify even more underserved youngsters if the PSA met some of the MI-specific conditions.

However, even without meeting the MI-related conditions, the PSA has still roughly doubled the identification rate of minority youth. Furthermore, meeting the general conditions -- as the PSA is likely to do in the next year or two -- means that it will be reasonable to infer judgments about students based on their performance on the PSA. Meeting these conditions will allow Charlotte to argue that it has doubled the number of underserved youngsters in its gifted program using a reasonably sound assessment.

Along with meeting these general criteria, the designers will still need to demonstrate that the PSA is a valid instrument. Toward this end, Romanoff is beginning to construct case studies of several identified students. To validate that the PSA actually detects youngsters who are or will be gifted will require evidence from many more sources. However, we are still awaiting such validation from standardized measures (see Chapter 1). In the meantime, the PSA is making high level curriculum more equitably available than traditional measures. Attempting to meet the conditions associated with MI may further increase the identification rate of underserved youngsters. In the final chapter, I will explore whether this is a step that the PSA is likely to travel given the dynamics of the district.

1. At second grade, when the district formally assesses youngsters for the gifted program, S.T.A.R.T. students were 50 percent more likely than those in a control group to be referred for gifted assessment. They were actually identified at 2.5 times the rate of the control group (Charlotte-Mecklenburg Schools, 1994a).

2. Along with an emphasis on problem solving akin to Maker's (see Chapter 2), this definition also appears to be influenced by Renzulli's three-ring conception of giftedness. (See Chapter 1.)

3. Murphy was superintendent from 1991 through the academic year of my visit (1995-1996). He was succeeded by Dr. Eric Smith.

4. Udall became moved from her position as coordinator of the Program for the Gifted to coordinating director of curriculum in 1995. In 1996, she became the assistant superintendant for curriculum.

5. This state policy will be superceded in the spring of 1998 by identification policies to be established by every local school district.

6. Sternberg's more recent ideas have influenced the PSA designers. For example, the designers include tasks that draw on his notions of practical, analytical, and creative intelligence.

7. In 1996-1997, the Program for the Gifted opted to devote more resources to the enrichment of the Challenge Team's classrooms. Thus, rather than evaluating all the students, children in "low-pop" schools receive several more months of pre-assessment lessons which helps them gain skills needed to perform well on the PSA.

8. This excludes the schools in which there are gifted magnets, and in which the population of identified gifted youngsters (at grades 3 and above) exceeds 40 percent.

9. Obviously, the results from these two schools yield questions about where in Charlotte-Mecklenburg the minority students are actually being identified. It may not be at the farther ends of the economic and integration continua, which these two schools represent. Instead, the results from the PSA in these two schools suggests the possibility that segregation in gifted education continues among the more segregated schools, with increasing identification rates for gifted African American and poor youngsters logically left to those schools that are more balanced racially and economically.

10. A numeric scoring system for writing used throughout Charlotte maps onto the PSA scores in approximately this way. A 1 is "probably not evident"; 2 is "probably evident"; 3 is evident or strongly evident; 4 is strongly evident or always evident.

Chapter 4
MONTGOMERY COUNTY'S EARLY CHILDHOOD GIFTED MODEL PROGRAM: A VISITOR

INTRODUCTION

In this chapter, I describe and analyze the Early Childhood Gifted Model Program, an effort to use MI to increase the identification of underserved gifted students in Montgomery County, Maryland. The work of the Model Program is quite different from that reported in the previous two chapters. Unlike the PSA and DISCOVER, the Model Program did not devise, or rely on, new and discrete assessment tasks. Instead, identification was supposed to draw upon teachers' efforts to elicit and develop their students' intelligences in the classroom curriculum and upon their observations of students in the classroom.

Unlike the efforts described in the preceding two chapters, the Model Program in Montgomery County Public Schools ("MCPS") was begun in a single school, Montgomery Knolls, rather than on a broader scale. After considering the theoretical and historical foundations of the Model Program, I describe efforts undertaken by Montgomery Knolls' teachers and staff to enhance identification. Following this, I analyze these efforts against the five general conditions to understand whether changes in identification can be linked to the Model Program's practices. Then I analyze the work against the MI-specific conditions to understand whether changes in identification can be linked to MI. (See Chapter 1 for a discussion of these eight conditions.) In the final chapter, I consider why these practices remained confined to Montgomery Knolls, despite the stated aim of the program to develop educational and identification practices for underserved youth that could be disseminated into the metropolitan area and state (MCPS, 1989).

THEORETICAL AND PROGRAMMATIC BASES FOR THE MODEL PROGRAM

Montgomery County Public Schools received two Javits grants which drew on MI to identify underrepresented youngsters for gifted programs. The first grant, awarded in early 1990, supported work in Montgomery Knolls, a pre-K-2 elementary school, to develop the Early Childhood Gifted Model Program. The second Javits grant was entitled Multiple Intelligences: A Framework for Student and Teacher Change (U.S. Department of Education, 1994) and was awarded in 1993. This grant was to continue the work at Montgomery Knolls and extend it to Pine Crest Elementary, the school into which Montgomery Knolls students articulate for grades 3-6. The second grant supported the program through December 1995, the time of my visit.

The aim of the effort at Montgomery Knolls was to build a model program to nurture the strengths of three groups that are traditionally underrepresented in programs for the gifted and talented (Montgomery County Public Schools, 1989; Starnes, n.d.). These are economically disadvantaged students, those with limited English proficiency, and gifted youngsters with learning disabilities ("GT/LD"). The Model Program included instruction and curriculum for these underserved youngsters, as well as "a process for identifying these students" (MCPS, 1989).

The Model Program built on a number of theoretical and programmatic foundations. The first of these is Montgomery County's existing Program of Assessment, Diagnosis, and Instruction or "PADI" (MCPS, 1989). This is a program aimed at developing the ability of underserved youngsters, especially minority youth (Gregory, Starnes, & Blaylock, 1990). PADI was first implemented in two elementary schools in 1981. By 1995 it was in 30 schools that have a higher concentration of poor and minority students. Like Charlotte's S.T.A.R.T., PADI selects youngsters for enriched classrooms with the aim of ultimately identifying more underserved gifted and talented youngsters.

Students are selected for PADI classes using a battery of seven diverse measures that minimize language demands and that have been deemed appropriate for use with minority students (Johnson, Starnes, Gregory, & Blaylock, 1985). Youngsters selected via this battery receive half- or whole-day enriched, interdisciplinary instruction. The curriculum emphasizes science and social studies and is taught by teachers with special PADI training. From these classes, approximately 25 to 30 percent of children are selected to participate in gifted and talented programs in the County (MCPS, n.d.).

A second important foundation for the Model Program was the existing philosophy of "identification through teaching" (MCPS, 1989; Starnes, n.d.). That is, data about youngsters' abilities are supposed to be gathered in the course of classroom teaching and observing, rather than gathered only in discrete testing activities. Waveline Starnes, the county's Director of Gifted and Talented Programs and Magnet Programs during most of the Model Program, described identification through teaching this way:

> What it is is that you notice sparks or indications of thinking ability.... You notice this ability -- Gee, you didn't even think the kid had it, and there: He solved that musical problem! Or you saw him solve this spatial problem that everybody in the room was trying to figure out how to do. You would never be convinced by that one indication. But teachers have drawers [for each student]. Well, stored in my head would be now: Robert's answer to that question was really unusual. And it would cause me to do something differently in teaching to test out and gain confirmation. And you go back and forth between that's a good idea and

> that's not a good idea: he's bright; he's not bright. I'm not just confirming, is he bright? I am [also] confirming how is the best way to teach him.

The introduction of MI expanded the existing identification through teaching approach. To ensure that underserved students' abilities would not be overlooked in the Model Program, "The curriculum will be developed to reach each student's strength ... as described by Gardner" (MCPS, 1989, p. 6). MI provided a framework for designing "action based, hands-on activities" to engage, develop, and identify youngsters' abilities. Mongtomery County's 1992 proposal to the Javits Program states that the curriculum developed under the first Javits grant drew on all the intelligences "integrated with science and the arts" to create "an environment in which students demonstrate exceptional strengths that might otherwise have been masked" (MCPS, 1992, p. i).

In its first three years, the Model Program also used some Spectrum assessment activities. (See Chapter 1.) When a child's strengths were not demonstrated clearly in the course of hands-on curricular activities in the classroom, special grant-funded teachers occasionally administered Spectrum tasks to diagnose a child's proclivities. This information provided feedback to the classroom teacher who could then use it to develop curricular activities to enable identification through teaching (MCPS, 1992).

Finally, teachers and grant-funded staff at Montgomery Knolls developed the Checklist for Identifying Learning Strengths, or "MI Checklist." (See Appendix K.) All classroom teachers at Montgomery Knolls used the checklist to observe youngsters. Some also used it as a tool to plan and develop instruction. It thus linked MI and the identification through teaching approach. (The checklist is discussed more fully in the Description of the MI-Influenced Identification Practices.)

HISTORY AND CONTEXT OF THE MODEL PROGRAM

Montgomery County Public Schools form a county-wide district, of 495 square miles, located just north of Washington, D.C. The district has 123 elementary schools, 29 middle schools, 21 high schools, and six special or alternative schools. These are organized into 21 geographically based feeder patterns or "clusters," each named for the high school into which the lower schools articulate. The district includes 120,000 students, and 7930 professional staff, among these 6933 teachers. The overall school population is 19.3 percent African American; 12.5 percent Asian, 12 percent Hispanic; and 55.8 percent White. Just over 21 percent of the county's students receive free or reduced meals, and 6.3 percent are enrolled in ESOL (English for Speakers of Other Languages) (MCPS, 1996a).

Although the county is considered affluent with well supported schools (Eaton, 1996), there has been a rapid increase in minority and poor youngsters in the past two decades. In 1978, the non-white student population was 18 percent. By the mid-1980s it was about 29 percent (Johnson, Starnes, Gregory, & Blaylock, 1985). In 1995-96, 44 percent were non-white (Eaton, 1996; MCPS, 1996a).

With the increase of minority students in the county has come an increase in segregated schools (Eaton, 1996). Some schools and some clusters, especially those in the northern and western part of the county, have few poor and minority students. For example, each of the three schools in Poolesville Cluster has between 88-90 percent white students and between 5.1-7.6 African American. In contrast, in the southeastern part of the county, adjacent to the District of Columbia, some clusters are predominantly African

American and Hispanic, with many poor youngsters. In Blair Cluster, which houses the two schools that I visited, 32.17 percent of the students are African American, 10.97 percent Asian, 25.37 are Hispanic, and 31.19 white. Across the cluster's 13 schools, almost 47 percent of the students are on free and reduced lunch, more than twice the county's average (MCPS, 1996a).

Despite large differences in the proportions of poor and minority youngsters across the county's schools, Montgomery County, unlike Charlotte-Mecklenburg, has never had its school assignments challenged in court. In 1975, the county adopted a voluntary desegregation policy. The policy sought to achieve desegregation mostly via magnet programs placed in Blair cluster schools. Eaton (1996) contends these policies and programs have proved largely ineffective in reducing racial or economic imbalances. Starnes argues that this perspective ignores the trend in the mid-1970s toward complete abandonment of white families from the area in question, a trend she believes the magnets prevented.

What is undisputed is that Blair cluster has many more poor and minority youngsters than affluent clusters to the west and north. At the present time, the district is not seeking to address such imbalances, but rather to limit their spread and to improve the quality of education within schools as currently configured (Eaton, 1996). Given this policy, along with describing and analyzing the county's Javits-funded work, this chapter considers the extent to which the Javits funding made a difference for those in the segregated schools in Montgomery County.

EARLIER IDENTIFICATION PRACTICES FOR GIFTED AND TALENTED PROGRAMS

Changes in Montgomery County's identification practices began in the late 1970s. Alongside the increase in minority and poor students, came a "growing concern about the under-representation of minority students in MCPS programs for gifted and talented students" (Johnson, Starnes, Gregory, & Blaylock, 1985, p. 417). While the proportion of minority students had risen in the district, there was not a corresponding change in the proportion of minority students identified as gifted and talented (Johnson, Starnes, Gregory, & Blaylock, 1985).

Until the late 1970s the county had relied on a two-stage identification process. In the first, or "global," stage all youngsters were screened largely via teacher recommendations and traditional standardized instruments. This initial screening was used to select a smaller pool of youngsters for "specific" screening. Only this smaller pool of youngsters was assessed with what Donnelly Gregory, the Coordinator of PADI, called, "the good stuff": the Raven's and other measures that are generally better at identifying minority youngsters. Given that global screening was weak in such measures (see Chapter 1), few poor and minority students were ultimately identified.

In response to this underrepresentation, Montgomery County initiated its Project to Minimize Socioeconomic and Cultural Barriers in the Education of Gifted and Talented Students in December 1980. The project sought to enhance African American and Hispanic students' access to programs for the gifted and talented through staff development, the PADI program described above, and revision of the identification process.

The county's revised "Procedures for Selection of Elementary Students to Participate in Gifted and Talented Programs" (MCPS, 1987a) is a multi-stage process that is still in use. Typically the process begins in second grade, and youngsters can be reevaluated annually. In the revised global stage, information is collected about all students from a variety of what the county refers to as "subjective" and "objective" sources (MCPS, 1987a). In the subjective arena are nominations from teachers, parents, adults in the surrounding community, and students themselves. In the objective category are scores from the Raven Progressive Matrices test, and sometimes other standardized tests. Any student with two or more of these objective or subjective indicators participates in the revised specific screening. Any minority youngsters with evidence from one indicator must also participate in specific screening (MCPS, 1987a).

The specific screening also includes subjective and objective measures. The subjective measures are peer nominations and two teacher checklists -- the Renzulli-Hartman Scales and Renzulli-Smith Early Childhood Checklist/Revised (various instruments used in the revised screening are listed in Appendix I). The Renzulli-Hartman asks teachers to note how often they have observed 8-10 behaviors that appear under three categories: "Learning Characteristics," "Motivational Characteristics," and "Creativity Characteristics." The four-point rating scale includes: "Seldom or never," "occasionally," "considerably," and "almost always." The Renzulli-Smith asks teachers to note how often they have observed 15 behaviors on a four-point scale: "seldom or never" to "always." Among these are "Displays unusual talent in music, drawing, rhythm, or other art forms," "Shows keen observation and retention of information about things he/she has observed," "Uses advanced vocabulary appropriately." In the objective realm are additional standardized tests, including the Test of Cognitive Skills. This is a standardized group test that presents figural puzzles and pictures. It yields scores in four categores: memory, analogies, sequence, and verbal reasoning. Finally, specific screening can draw on "other performance data," including additional standardized tests, information about reading and math levels, acceleration to a higher grade, or "other outstanding performance" (MCPS, 1987b, p. 53).

In some schools, including Montgomery Knolls, the global and specific screening are combined at the discretion of the staff. In such cases, the classroom teachers complete the Kough/DeHaan Checklist. On this checklist, teachers fill in the names of their students who can be described by various behaviors, including "Learns rapidly and easily," "Is independent, individualistic, self-sufficient," and "produces original products or ideas." After this, the combined screening requires teachers to complete the Renzulli-Hartman for any student who meets at least one other indicator, even if that indicator is not a standardized test (MCPS, 1987a, 1987b).

When all this information is scored, it is compiled onto a grid listing the indicators/measures across the top, and the names of students down the left side (See Appendix H). Then, at each school a screening committee is formed of the principal and some staff at each school. This committee convenes to review and discuss the information about each child. The county's formal identification guidelines note that "the grid sheets will enable the committee to identify three groups of students" (MCPS, 1987a, p. 8). In Group I are those who clearly meet the formal identification criteria: strong scores on three or more indicators. In Group III are students who reveal "few if any indicators." The committee therefore need not devote much time to considering their eligibility for gifted and talented services. Group II students, "show one or two indicators" from the specific screening measures. These students need to be discussed individually by members of the screening committee (MCPS, 1987a, p. 8).

The county has compiled a reference chart to help interpret scores for each of the indicators at different grade levels (Appendix I). At the same time, each school's screening committee is empowered by the county's procedures to exercise professional judgment in interpreting information from the grid sheet:

> ...these indicators [should] be regarded as tools for decision making, subject to professional interpretation rather than rigid cutoffs. For example, no child should be excluded for missing an indicator by one point. The standard error of measure of group test scores may make them an underestimation of the child's true score. This is frequently the case for black and Hispanic students. Such scores for some students can be corroborated by other information that supports their high performance and ability. The school committee may then feel that the information available is sufficient for them to make the decision that the student needs differentiated programming even though test data do not seem to support the decision (MCPS, 1987a, p. 8).

Teachers are encouraged by the official county procedures to advocate for students (MCPS, 1987a). In such cases, the screening committee can examine students' performances or products, hold structured interviews, or seek additional information (MCPS, 1987a, p. 9). From interview data, it was clear that teachers in the screening committee at Montgomery Knolls did act in this way (See Conditions 2 and 4, below.)

EDUCATION OF THE GIFTED IN MONTGOMERY COUNTY

What does it mean to be gifted in Montgomery County? The formal meaning runs parallel to recent federal definitions (see Chapter 1):

> (1) Children and youth with outstanding talent who perform or show the potential for performing at high levels of accomplishment when compared with others of their age, experience, or environment. (These talents are

present in children and youth from all cultural groups, across all economic strata, and in all areas of human endeavor.)

(2) Children and youth who exhibit high performance capability in intellectual, creative, and/or artistic areas, possess an unusual leadership capacity, or excel in specific academic fields. (They require services or activities that may go beyond those ordinarily provided by the schools.) (MCPS, 1996b, p. 2).

The pragmatic meaning of identification is much less clear. At the present time, programming is, as one staffer said, "a mish mash." Over the last seven years, Montgomery County has undergone many changes in services for gifted and talented youngsters. In January 1990, the county had 13 teachers of the gifted specifically assigned to help administrators and teachers develop curriculum for identified youngsters as well as other students. In 1991, there was a sharp budget cutback leaving only three teachers to serve as resources for the entire county.

At that point, given both budget cuts and concerns for equity, regular classroom teachers were supposed to differentiate instruction for the whole range of learners. In addition, individual schools can still organize programming for gifted youngsters by clustering advanced youngsters at various points in the day. Given this, "Every school does gifted and talented differently," according to Pine Crest's principal, Pam Sobel.

Despite budget cuts, there are still important opportunities specifically reserved for elementary students who are formally identified. One of these is the ability to attend two elementary gifted magnets, both in Blair Cluster. A second is the possibility of attending one of the four centers for the highly (or as one staffer jested, "severely") gifted. The latter are highly competitive, full-day programs that serve a total of 200 fourth graders and 200 fifth graders selected from throughout the county.

This kind of challenging curriculum in elementary school helps prepare youngsters for magnet programs in middle school, for which students throughout the county compete, and for the International Baccalaureate Program in middle school and high school. The magnets programs have drawn white students into Blair cluster schools, while minority students from the wider school population are infrequently selected to participate (Eaton, 1996).[1]

In short, identification still brings substantial benefits: access to centers for the highly gifted and to elementary gifted magnets. This in turn better prepares identified youngsters for competitive enriched programming in middle and high school. Furthermore, recent county documents indicate a rethinking of the loosely structured programming currently offered in most elementary schools. The county is now advocating "systematically provided" services to this population (MCPS, 1996b, p. 1; MCPS, 1996c). If such plans are put into effect, the stakes associated with identification could well increase.

THE IMPLEMENTATION OF THE JAVITS PROGRAM AND ITS IMPACT

While PADI and the 1987 revision of identification procedures improved the representation of underserved youth, some in the county asserted that more needed to be done to identify such youngsters. A 1988 Report of the Superintendent's Advisory Committee on the Education of the Gifted and Talented called for developing comprehensive parent outreach for minority parents, recruiting minority staff to the gifted and talented program, expanding PADI to some 12 additional schools which had large minority populations, providing PADI staff development for teachers not in PADI schools, and increasing the number of Hispanic gifted and talented students (MCPS, 1989, Appendix F). Such students were served by PADI, but moved into gifted programs at a rate lower than African American or poor youngsters (MCPS, 1989, Appendix F).

An opportunity to act on this need came through in 1989 with the first round of requests for proposals from the Javits Program. The county's proposal expressed "the need for a more comprehensive program and additional strategies to serve limited English proficient and Hispanic students" (MCPS, 1989, p. 2). Despite the revised identification process and PADI, Starnes asserted that a "verbal veil" obscured the strengths of students whose language capabilities were not obvious to teachers. She felt that the verbal veil especially applied to learning disabled students, poor youngsters, and those with limited English proficiency.

During the drafting of the county's first Javits proposal, Starnes and her assistant, Deborah Leibowitz, sought to identify a school site in which the proposal ideas would likely bear fruit. Within a short time, they settled on Montgomery Knolls. Leibowitz became the Program Specialist for the Model Program, helping to orchestrate the program's development at the school.

Montgomery Knolls is a pre-K-2 elementary school located in Silver Spring, Maryland. In 1995-1996, the school had about 400 students, 39 percent African American, 28 percent white, 19 percent Hispanic; 13 percent Asian.[2] Almost 48 percent of the students received free or reduced lunch and nearly 9 percent are ESOL (MCPS, 1996a). The school is situated in a neighborhood that seems suburban, with mostly small, brick, single family homes dating from the 1950s, surrounded by yards and shaded by tall trees. Yet, about a mile away are low-rise brick projects, in, at most, modest repair.

Montgomery Knolls was built in 1952 and rehabbed in 1989. It is pleasant inside: Large windows look out onto the treed landscape. Big classrooms are each equipped with at least two computers, manipulatives, and bulletin boards full of students' work. The staff in 1995-1996 included 12 classroom teachers. There were music, art, and PE teachers (.9 FTE each), as well as specialists for reading, media, the resource room, computer magnet curriculum, and other areas. There are also several paid classroom assistants (MCPS, 1996a).

When the Model Program began, Montgomery Knolls was blessed with an exceptional principal, Pamela Prue, an African American woman who was the Montgomery County recipient of the Washington Post's Distinguished Principal Award in 1993. Virtually everyone who spoke of Prue praised her ability to engage teachers in the process of educating children and improving their own practice. In addition, the teachers at the time of Prue's leadership were mostly extremely dedicated veterans (Krechevsky, 1992). Beyond these vital resources, Montgomery Knolls was a county wide computer magnet school, had a PADI Program, an all day kindergarten, Chapter I funds, the Comprehensive School Mathematics Program, a mentoring program, and several years' experience with whole language instruction (Krechevsky, 1992; MCPS, 1996a).

The second Javits grant was partly to enable MI-influenced practices to be implemented at Pine Crest, the elementary school into which Montgomery Knolls students feed. While MI has influenced teachers at Pine Crest, its implementation was far slower and more problematic than at Montgomery Knolls. In the county's evaluation of the grant, teachers' own ratings of their understanding and use of MI were consistently lower at Pine Crest than Montgomery Knolls (MCPS, 1996d, Appendices C and E). Pam

Sobel, Pine Crest's principal, noted several impediments: The staff really had much less training than the those at the beginning of Montgomery Knolls Javits Program. Teachers also had not yet grasped the importance or meaning of multiple intelligences when they were asked to use the MI checklist. In addition, they perceived MI and the county-wide assessments that begin in third grade as partly in conflict. Sobel asserted that only at the grant's end in 1995 were teachers "at a stage where they are starting to understand and process" MI (MCPS, 1996d, Appendix N, p. 3). Given that implementation of MI was weak at Pine Crest, it makes little sense to attribute any changes in identification there to the theory. Thus, this chapter focuses on whether outcomes realized at Montgomery Knolls can be associated with the practices actually adopted there.

To implement the grant at Montgomery Knolls, Starnes and Leibowitz organized two "think tanks" just prior to receiving the first Javits grant: One of the think tanks was a series of meetings for county staff "who had anything to say about young children." This included county experts concerned with special education, ESOL, all areas of the curriculum, gifted and talented, Head Start, as well as school psychologists and representatives of the Department of Educational Accountability. These people were asked to brainstorm a question: What would you see if the roof were removed from the ideal school for young children? Leibowitz collected these answers, organized them, and had the experts review and refine them over several sessions. That same question was posed in the second, day-long think tank, composed of the staff and teachers of Montgomery Knolls.

Leibowitz reported that across the two groups "essentially the key elements were the same." These elements were ultimately organized into a "tapestry" to represent ideas and practices that supported the Model Program's aims. "The tapestry of program strands became a framework that helped teachers find learning gifts frequently masked in some children" (Leibowitz & Starnes, 1993, p. 30). The strands included many different ideas, including "active learning," "constructivist learning," "[a] problem-solving focus," and "community of learners." According to Leibowitz, "MI was the undergirding or the foundation...." After many revisions, the tapestry was represented with MI eventually running lengthwise across the top, like a rod from which the entire tapestry hung (See Appendix J).

These program strands were elaborated and fitted to the ongoing needs of Montgomery Knolls by a monthly meeting of the "grant council." This group included the grant staff, the principal, teachers representing each grade, plus teachers of reading, ESOL, and sometimes special area teachers and outside consultants. Leibowitz said these were the "people who could provide the widest picture of what was going on from their own constituencies and bring the widest set of problems. In other words, [they would] be able to identify where there were issues that needed to be addressed." Between the think tank at the school and the grant council's monthly meetings, the majority of teachers at Montgomery Knolls felt a sense of ownership and investment in developing the project.

While there was clearly a great deal of effort made to implement MI, the changes in identification rates that might be associated with this effort are less clear cut. As Table 4.1 reveals, in the spring of 1989, the year before the implementation of Javits-funded work at Montgomery Knolls, 23 percent of the second graders were identified. In the next two years, the identification rate was nearly identical, perhaps because practices were not yet in place long enough to stimulate changes in identification.[3]

In the following three years (1992-1994), the percentage of identified youngsters more than doubled. Thus, by 1994, 51 percent of the second graders were identified as gifted and talented. This figure dropped down to 31 percent in the spring of 1995, the last year of Javits funding.

TABLE 4.1: Second Grade Students Identified as Gifted/Talented at Montgomery Knolls (MCPS, 1996d, p. 15).

Spring of Year:	1988	1989	1990	1991	1992	1993	1994	1995
Total 2nd Grade	91	81	73	94	89	117	77	91
Number GT	25	19	16	22	41	56	39	28
Percent GT	27	23	22	23	46	48	51	31
Number of GT:								
African American	5	7	6	9	8	17	18	6
Asian American	3	4	2	1	9	8	7	4
Hispanic	1	2	1	1	4	3	3	3
White	16	16	7	11	20	28	11	15
Percent of GT:								
African American	29	37	38	41	20	30	46	41
Asian American	12	21	13	5	22	14.5	18	14
Hispanic	4	11	6	5	10	5.5	8	11
White	64	32	44	50	49	50	28	54

Though percentages of identified second graders increased dramatically in 1992-1994, Table 4.1 also shows that there were not commensurate changes in the proportions of traditionally underrepresented students. With the exception of 1994, the percent of African American students identified was not markedly higher than it had been in the two years preceding the grant. The percent of Hispanic youngsters identified during the six years of the grant (1990-1995) was between 6 and 11 percent. This fell within the range established by the two years preceding the grant. The same general pattern holds true for white youngsters: In 1988 and 1989 white youngsters made up, respectively 64 percent and 32 percent of those identified. After the grant, the identification rate for white children held to within those boundaries, with the exception of 1994 when it fell sharply (MCPS, 1996d).

In the context of Montgomery Knolls' population (c. 40 percent African American; 19 percent Hispanic; 28 percent white; 13 percent Asian), during the grant years, African American youngsters were sometimes identified as gifted at a rate proportionate to their presence in the school population. Hispanic youngsters were consistently underrepresented, and white students were overrepresented, except in 1994.[4]

Because the target populations for Montgomery's Model Program also included GT/LD youngsters, it is possible that the MI intervention yielded higher percentages of these children. In a continuation application, the program staff claimed that 17 percent (n=7) of the 41 second graders identified at Montgomery Knolls in 1992 were "possibly learning disabled" (MCPS, 1993, Appendix D, p. 1). However, Brian Bartels, Montgomery Knolls' school psychologist and later the grant-funded psychologist who analyzed much of the Model Program's data, said that additional quantitative studies of identification of LD/GT students were not undertaken. There were only some "very subjective" case studies. In addition, it is possible that more poor youngsters were identified. Staff who sat in on screening youngsters for gifted strongly believe this was the case. Unfortunately, there are no quantitative data to help support this claim.

In sum, at Montgomery Knolls early reports indicated large increases in the overall school population identified as gifted (Leibowitz & Starnes, 1993). While this claim holds for several years of the Model Program, the proportion of Hispanic and

African American youngsters identified is not markedly different than existed prior to the grant.

DESCRIPTION OF THE MI-INFLUENCED IDENTIFICATION PRACTICES

While there was not a noticeable increase in the proportion of minority students identified in the Model Program, it is worth looking at the practices developed there for at least two reasons. First, for three years (1992-1994), the number of identified students did rise steeply. Something happened at Montgomery Knolls that made more children perform in an advanced way, and/or made their teachers perceive them as doing so. Second, by examining the practices used, it becomes possible to hypothesize about why there was little change in the proportion of underserved youth who were identified. If the practices themselves do not suggest some deficiencies, then it is reasonable to look for other causes.

As noted earlier, Montgomery's Model Program did not employ a distinct set of assessment tasks inspired by MI theory. Identification for gifted and talented continued to be based on the county's 1987 revised global and specific screening described earlier. What did change with the grant was the introduction of instructional practices aimed at developing and recognizing children's strengths. These practices provided an expanded basis for "identification through teaching." While the program tapestry lists over 20 elements of the Model Program, the practices that teachers and designers highlighted are described below.

Whole School Treatment

Unlike PADI's special nurturing classes, which were available to youngsters selected mostly via PADI battery scores, the Montgomery Knolls' Model Program was school wide. From the think tank came a notion, as Leibowitz put it, that "when you deal with a regular school, especially a population such as Montgomery Knolls', you realize all children fall into the category of possibly gifted, but definitely underserved, in the normal run of things." Thus, the children in all K-2 classrooms were provided opportunities to draw on and develop their diverse intelligences.

The school-wide approach was not only extensive but intensive: The grant staff believed that it is difficult to discover children's diverse strengths unless students are given opportunities throughout the school day to display and develop them. Thus, "to seek and nurture these special, but sometimes hidden intelligences in each child, the program is in place for every child in every classroom, in every special center, and in every learning space in the school" (Leibowitz & Starnes, 1993, p. 30).

The school-wide involvement was fostered by mandating the participation of all teachers, though participation was understood to be at each teacher's own pace. Teachers at Montgomery Knolls were not asked to reorganize their classrooms from scratch. Instead, Prue encouraged teachers to review their existing curricula through an MI lens and to think about how their curriculum offered opportunities for children to engage their diverse intelligences. One teacher recounted sitting in the middle of the different centers in her classroom and actually working through this exercise. To further teachers' involvement, Pam Prue asked teachers to set their own goals for using the theory, and these were among the goals she used in her annual teacher evaluations.

Teachers and administrators indicate that under Prue's leadership, the great majority of the staff invested effort in rethinking and reorganizing their approaches to enable children's multiple intelligences to develop. In 1992, one teacher reported that you could still find cars in the parking lot at 5:30 on a Friday night (Krechevsky, 1992). The county's final report to Javits noted, "After two or three years of implementation at Montgomery Knolls there were a few teachers ... who remained very skeptical of the value and originality of using MI in the classroom. Many of that group of dissenters from the primary school are now among the greatest advocates of using MI (MCPS, 1996d, p. 2). The school wide influence was highlighted by Prue. She reported that with the introduction of MI, "I saw a spark in teachers that were [already] great teachers.... I mean, we were just stamping in the halls, everywhere you would go in that building, everybody was just turned on to the idea that, boy! We were seeing kids differently!"

Curriculum

Once the grant had been received and the think tanks organized, Leibowitz claimed that "it was obvious to everybody, it became blatantly obvious to the most casual observer, that you had to provide opportunities to children to explore each of the intelligences in order to find out which strengths the children had."

As Prue said:

> As we began to explore the theory, we recognized that there were five other areas of intelligence or thinking in children [besides linguistic and logical-mathematical]. Then there was certainly the strong recognition: well, we're going to have to transform these learning environments and provide these kinds of opportunities, so that indeed we'll get to see kids thinking using those strengths.

Carol Hylton, who worked at Montgomery Knolls as a second grade teacher, and then became the Javits Grant teacher trainer for both Montgomery Knolls and Pine Crest, expressed the same view:

> How can you see spatial abilities, bodily-kinesthetic abilities, musical abilities, if you never provide opportunities for kids to use them? Or to develop them? If they don't have chances to use music or to use their spatial abilities in class, how are you going to know if they're able in that way?

The curriculum that evolved to identify and build upon children's multiple intelligences had several complementary and overlapping elements. These included:

<u>Authentic Activity/Learning Centers</u>: At the kindergarten level, teachers had "play centers" in place prior to drawing on MI. The centers included a variety of materials conducive to art, construction, imaginative play, and other areas. To reformulate these play centers through an MI lens, the teachers initially expanded them so that there would be "one for each of the intelligences." This enabled the teachers to gain some understanding of behaviors associated with different intelligences. However, after some time doing this, they came to see intelligence-focused centers as a naive approach. They wanted the centers to reflect real-world activities, such as drama, construction, and sports, which draw on combinations of the intelligences, as Gardner (1983) has argued.

With this realization, the teachers reworked the centers to reflect "authentic activities." In a visit to a kindergarten classroom, some of the centers I observed included music, computer, movement.(including a basketball hoop and sponge basketballs, and equipment on which to balance at low altitude), art, construction (stocked with Lego® blocks and other construction materials), and drama (with a puppet theatre and backdrops).

Exploratory Centers: After reworking their centers, the teachers wanted to push beyond providing materials that allowed children to explore authentic activities. Karen Bulman, a kindergarten teacher, reported that teachers decided "to set up what we called exploratory centers, where there would actually be some problems that the children would have to solve." The problems that were laid out were open-ended, such as an assignment to develop a story for a puppet play. They could be conducted in small groups or individually, did not draw on adult participation, and ended in some oral reporting of what was done or learned. Bulman relied on a rotation system enabling children both to choose exploratory centers and to visit all the different centers. This provided information about which centers children preferred and to observe children interacting and problem solving across a variety of media and materials. Because of this system, one teacher reported, "We could actually do some assessing through those centers."

I observed Bulman assessing center activities by jotting notes on Post-Its® that were spurred by her observations of children as they worked. She reported using the notes to remind her of things "that I need to work on with a child." In addition, she keeps the MI checklist on her clipboard, "So, if I'm focusing on one or two children, I can actually just make notes right up here on the top that I might really want to be aware of. Especially on the child that I have not seen any real spark...." Some of the other teachers used "insight cards," half-sheets of paper on which observations or insights about a child could be made and put into the child's folder for the teacher to use in planning instruction or assessment activities.

Thematic and project-based curriculum: To provide a context for first and second graders to draw on diverse intelligences in their problem solving, and to help them acquire content, teachers relied on broad-based themes (Leibowitz & Starnes, 1993). Some of the themes developed during the Model Program included weather, water, construction, and change. Within themes were particular projects or units. In a unit on cowboys, Barbara Williams, a first grade teacher reported:

> We did "Home on the Range" and we did a little song called "I've Been Riding on the Range." They talked about herding doggies. We went into what does a cowboy do, once the range was established. You know, you can refer back to that song, interpreting what home on the range means, if you're out west. They made murals of the west and pictures. They tried to find ... pictures that would show the range. We made up a game when we went outside: ... They were all cattle, and it was played like free tag. And we took them outside and we said the big field is the range, and the cowboys job is to 'round you up.' And he's going to round you up by tagging you... So you know all those activities were taught at the same time to develop the concept of a range and the cowboy's job.

Carol Hylton explained that a science unit on butterflies:

> allowed multiple access points over several weeks [involving student] data gathering and representing understanding. So I'd have a lot of art material available for children to use in the process of their developing understanding of butterflies and of metamorphosis. I would specifically structure some things that were B-K [i.e., bodily kinesthetic]. I would specifically structure some things that were logical-mathematical. And I would do a lot of observing.

Hylton went on to explain that she kept much of what she observed in her head, but she also used insight cards for "notes ... about a specific kid or about a kids' interacting with some piece of content or some event." Thus, information about students' abilities was collected while allowing students to draw on different strengths to develop their understanding of a topic.

Linguistic Links: MI helped teachers and administrators to understand that youngsters might be talented even if they didn't evince language strengths. However, actual identification of youngsters' gifts was difficult for teachers -- even after their initial round of MI training -- in "children who were not verbal" (Leibowitz & Starnes, 1993, p. 32).

In order both to make identification more likely, and out of concern for the development of children's language skills, teachers devised "linguistic links": These were activities requiring youngsters to use words to describe what they did in solving a problem or making a product. For example, as indicated above, exploratory activities in the kindergarten are followed by youngsters explaining to their classmates what they learned or made. For youngsters who are reticent, Bulman sometimes had children bring their object or stand near the exploratory activity so that they could both talk about and demonstrate what they did. Another example of a linguistic link is writing that followed construction work or art work. For example, in the butterfly unit, after children made a three-dimensional butterfly, Hylton had them do procedural writing, detailing the steps they used to build the butterfly, and how the butterfly itself worked.

A number of other elements of curriculum were important to the teachers, including "action based, hands-on curriculum," "science," and "the arts." The same sorts of patterns hold through these elements: children had many ways to engage topics. They were called upon to use language and develop mathematical understanding, but they were not limited to developing primarily these abilities, nor were they assessed or observed primarily for such abilities. The entire range of intelligences was valued by the teachers and administrators I spoke with about the Model Program. As Prue reported, teachers were "really recognizing and tapping these [diverse] strengths."

Assessment

To document and further develop youngsters' strengths, teachers employed several different assessment strategies. As described above, teachers documented observations in various formats, from Post-It® notes to "insight cards." However, by far the most prominent documentation associated with the Model Program was the Checklist for Identifying Learning Strengths or "MI Checklist."

The MI Checklist evolved over many drafts by teachers and grant personnel, and it drew on feedback from Project Spectrum staff. (See Chapter 1.) It is now a two-page document, with seven sections, one for each of the seven intelligences. In each section, there are between seven and 11 observable behaviors associated with that intelligence. (See Appendix K.) For example, for linguistic intelligence, behaviors include "Enjoys word play: chooses to memorize and recite poems, tongue twisters, puns, riddles, etc.," and "Talks through problems, explains solutions." For interpersonal intelligence, behaviors include "Eager participant in group activities;" and "Easily builds relationships with others." For each behavior, teachers write in a number from one to five, to indicate how often the behavior has been exhibited: "not observed," "occasionally observed," "usually observed," "almost always or always observed." A five designates "no opportunity to observe" the behavior. Each section is also given an overall rating on the 1-5 scale. On the second page of the checklist are six lines for teacher comments.

Montgomery Knolls' classroom teachers completed the MI Checklist for each child at least twice a year, once in the fall and once in the spring. The checklist was supposed to inform both instruction and identification (Starnes, n.d.). In fact, the checklist served many purposes. First, the checklist was to provide "the basis for a common language ... more or less a definition of the seven intelligences, with characteristics," according to Leibowitz. The checklist developers, largely teachers on the staff, asked themselves questions like "a young child who was bodily-kinesthetic, what would we see in school? What could a parent see? What could a stranger see? ... That was the kind of thing that I used as an example to get people working on it."

Once formalized, the checklist served as a tool for many teachers to build in opportunities into their classroom for youngsters to demonstrate the behaviors on the checklist. This usage is illustrated by Bulman's referring to, and jotting notes on, the checklist as she walked around the classroom. Hylton said, "it was a trigger to kind of see how a kid is, and what you could do" to find out more about a child. "Hopefully, you were using it instructionally. Hopefully, you were planning for kids based on it." Jean Barton, the first psychologist for the Model Program and the Program Specialist after Leibowitz, said the checklist provided teachers with feedback. If they did not see a child having a strength in one of the intelligences they had to ask:

> 'Am I not seeing it because I'm not providing the environment through which I would see it? Am I not seeing it because it isn't there [in the child]...?' I think one of the things that we found out was that, for teachers who really internalized the Gardner model, the checklist was very much functioning as a teacher instructional planning and assessment tool....

Leibowitz, who first suggested the development of a checklist, believed that this "teaching piece" was its primary purpose. Yet, the checklist also aided teachers in a variety of assessment tasks. For instance, teachers were encouraged during the first grant to develop student portfolios "to help assess student progress and strengths in the various intelligences" (Starnes, n.d., p. 51). The checklist helped some teachers to organize student portfolios. Williams, the first grade teacher, said "We would use the checklist, and ... let's say, something [i.e., a behavior] that was under linguistic, we might collect the sample that showed that particular strength" for the portfolio. The checklist served "like a benchmark for things to look for."

Teachers also used the checklist in their conferences and meetings with parents. Prue said that "the lightbulb went on: you know, that we need to be writing this down and sharing with parents what we've found out thus far about what their youngsters' strengths are." Teachers resisted actually sharing the checklist with parents, "because they thought the parents would see it as, 'Well, why isn't he [my child] linguistic.' Or, 'How come you didn't find that he's spatial?'" Instead, teachers shared the checklist information informally with parents, especially to discuss children's strengths.

In Montgomery Knolls, the checklist also acted as a framework for thinking about children's talents and strengths for the purpose of screening for gifted and talented identification. Prue reported "that was the tool that we were using quite a bit to really identify -- to observe and identify these strengths." However, as discussed in the following section, the checklist was never a formal tool for identification.

The Screening Committee[5]

While the Office of Enriched and Innovative Instruction in Montgomery County advocates "identification through teaching" (Starnes, n.d.), and this philosophy guides some teachers in their efforts to develop youngsters' strengths, the actual designation of students as gifted and talented occurs in screening committees. These committees are organized at each elementary school and include the principal, counselor, and a subgroup of teachers. At Montgomery Knolls during the two Javits grants, the screening committee included the second grade teachers, since formal identification entails students at that grade. It also included one representative from each of the other grades, the school psychologist, typically the art or music teacher, and any other teachers who wished to participate. The committee usually met two or three times a year.

The bases for identification by screening committee members throughout Montgomery County are the multiple objective and subjective measures described earlier. These are gathered via the global and specific screening, or the combined screening, of students. *Neither MI, nor the MI checklist, was ever a formal basis for identifying youngsters for gifted education at Montgomery Knolls, Pine Crest, or elsewhere in Montgomery County.*

As Leibowitz put it:

> I don't think we altered the identification process for gifted and talented identification formally within Montgomery County Public Schools. Never did change that.... [T]he way MI informed the identification process is that the teachers brought the profile of MIs with them [to the screening committee]. Not necessarily in the checklist, although it might have been the checklist. By then, they knew the checklist backwards and forward. They knew each child in their classroom. They didn't need the formal piece of paper. What they brought with them was evidence of the particular intelligences that the child used in solving problems and creating products, and in just general work in the day-to-day existence in a classroom.

In essence, MI informed teachers' and the principals' professional judgment. As noted earlier, the county's official guidelines for identification call on teachers to exercise such judgment and to advocate for students (MCPS, 1987a).

That MI did shape some screening committee deliberations, at least during Prue's tenure at Montgomery Knolls, is clear from many teachers' and administrators' comments. Brian Bartels, the school psychologist said of the screening committee, "Before [MI] ...

the staff was kind of forced to look through the logical-mathematical and language lens."

The same sentiment was expressed later by Bulman:

> Before the grant actually came and we learned about the multiple intelligences, it [screening] was pretty much based on the test scores of the children. There was some teacher input. But I don't think we knew the kids as well as we do now. You know, we knew what they could do linguistically and logically, logical-mathematically, but we really hadn't explored those other areas. And now, after the grant came and we explored the multiple intelligences, all of that comes into play, and there is a lot of discussion.... Those areas [multiple intelligences] certainly came into play."
>
> Hylton said:
>
> When you came to the screening committees, the advocacy that I heard in the GT screening committee was based on teachers having seen kids differently and therefore rallying for individual children and their strengths -- diverse strengths: not strengths necessarily in linguistic and logical-mathematical areas, which might show up in some of the screening measures, but those more elusive ones that aren't normally tapped.... I heard my colleagues ... describe children in terms of their strengths. And they would use back-up data from what they had experienced with a child in class: 'But I saw him do this,' or 'I saw him do that.' 'Over the years, she did this,' and 'Do you remember when she was in kindergarten she wouldn't do this? But you're telling me now that she -- ?' ... And it was really very inspirational in a way: a lot of what the kids are able to do, teachers saw as indicators of possibility.
>
> Prue noted that:
>
> After the theory, and the Javits grant exploration, we would have these actual constructions come in to the [screening committee] table. And you just set 'em right here. You know, they're multidimensional, and you go: 'Whoah!' The teacher would provide the context for the creation. And you could just see the tremendous amount of thinking that these youngsters had [done] and the creativity that they had put into them.

Although teachers saw many youngsters differently after the introduction of MI, their ability to draw on this new information for the purposes of formally identifying children was limited. The school-wide opportunities afforded all children to demonstrate and develop their talents in classrooms was not paralleled in the county procedures that govern the screening committee. As described earlier, the county guidelines stipulated that identification of children in the top and bottom thirds was based on measures that appear on the grid sheet (see Appendix H). The opportunities to consider a youngster's multiple intelligences was mostly linked to the middle third or "Group II."

Hylton reported:

> As in most screening, there are the ones who are obviously yes, and there are the ones who are obviously no. And then there's the middle. And the middles are the ones that you are dealing with for the most part, because you're wrestling with whether or not they fit. Whether or not they need to be included in GT programming.

Leibowitz' comments confirm this:

> What they [teachers] were able to do, and what is legitimate as part of the formal identification process, is for those children where the data is hazy -- and that was a good third of the children -- ... we talked about them. And that's where the kind of teacher observation, and work samples, and evidence of problem-solving skills come into play.

Thus, for the most part, the existing county guidelines, determined identification. MI influenced the identification procedures of the screening committee mostly in the ambiguous cases of children in Group II. In these cases children's work was sometimes brought to the table and the discussions were memorable. However, the extent to which these powerful discussions occured was limited. Williams, the first-grade teacher, reported that MI entered the discussion only in a few cases during each meeting. The strengths in children that teachers saw emerge in their MI-influenced classrooms were not the strengths that could regularly be considered in the identification process.

ANALYSIS OF WHETHER INCREASED IDENTIFICATION CAN REASONABLY BE ASSOCIATED WITH THE MODEL PROGRAM'S PROCEDURES

In the previous two chapters, I relied on five general conditions to analyze whether it was reasonable to associate the increased identification of youngsters as gifted with the new assessment practices that each site developed. I also drew on three MI-specific conditions to understand if the assessment practices could reasonably be associated with MI. In Montgomery County, the formal assessment procedure was not altered. As one teacher said, "in terms of a certain set of activities, no, we didn't have 'em." Instead, classroom practices were put in place with Javits funding that may have influenced how teachers perceived and developed youngsters' abilities and the way teachers advocated for students at the screening committee. Thus, in the section that follows I attempt to analyze whether it is reasonable to associate these practices with increases in identification. When a condition is not met, I suggest ways to strengthen the approaches that were put into practice.

GENERAL CONDITIONS

Condition 1: Children Understand the Tasks

Since there weren't identification tasks at Montgomery Knolls, one way to think about this condition is: were the children enabled to understand the classroom experiences upon which "identification through teaching" was based?

A number of features emphasized at Montgomery Knolls might be said to help children develop and display understanding. For example, the curriculum drew on thematic units. These units lasted over several weeks, so that children could explore the content and become familiar with it. The units encouraged exploration through a variety of learning experiences, including hands-on approaches, music, art, movement, numbers, speaking, and writing. As the description of the cowboy unit illustrates, varied learning experiences, drawing on diverse intelligences, were combined in curricular units to enable youngsters "to develop the concept...."

Given that the units extended over time, incorporated reflection through linguistic links, and drew on diverse ways of representing and using information, it is possible that youngsters' understanding was enhanced (Gardner, 1991b; Perkins, 1995). However, confirming this possibility is difficult. There was no other control or "treatment" of youngsters in the school. Also, because no county-wide assessments are given before third grade, it is not possible to compare children's understanding or academic achievement at Montgomery Knolls with children at other schools.

Relative to the sorts of specific activities, materials, and directions employed in Charlotte and by DISCOVER III, in the Model Program there was likely greater variability in students' understanding: Understanding is variable in almost any heterogeneous classroom. My own observation of classrooms at both Montgomery Knolls and Pine Crest was that, outside the kindergarten (where children were completely absorbed in the different centers), there appeared to be a normal range in children's engagement and, likely, their understanding. A few here and there were upset about something unrelated to the lesson, were distracting other children, or were distracted by their classmates. Given such observations, alongside the descriptions of curriculum and efforts made to develop students' understanding of it, it is not possible to say youngsters understood the classroom experiences which influenced identification. It is only possible to say that efforts were made to meet this condition, though a good deal of variability in actual understanding remained.

Condition 2: Children are Encouraged to Do their Best Work

There were several features of the Model Program at Montgomery Knolls that did encourage students to do their best work. A number of these, mentioned above, relate to the curriculum. For example, curriculum units encompassed a range of experiences designed to engage children and develop their understanding. Furthermore, units lasted over several weeks. Engagement over time is a prerequisite for doing work well (e.g., Simon, 1979). Reflection, of the sort fostered by linguistic links, is another prerequisite for achieving best work (e.g., Perkins, 1995; Schön, 1983). Because the content was structured into integrated themes and units, students were helped to have a context for organizing the information being conveyed. This, too, fosters good work (Ceci, 1990; Resnick, 1987; Rogoff & Lave, 1984).

In addition, teachers reported that implementing MI helped them to evoke children's best work. Hylton said MI was what she needed "to address both their strengths and to create an environment that was more conducive for children to learn and to stretch within the class."

With the application of MI, teachers became better able to see the best work that children could do. According to Bulman, a kindergarten teacher:

> I guess we're just so much more aware of how the children work best, because we've offered them the opportunities now to show us. ... [Before the grant] we just hadn't set up experiences where they could choose to work alone or choose to work in a group. We were always telling them how to do it. And now that we've set up these [different] activities, you really see how they do their best.

The staff came to see certain behaviors, for example a girl pirouetting in the corridor, not as violations of deportment, but as clues about how to engage children and develop their abilities. Brian Bartels, the psychologist, noted:

> There was so much going on with art and music, with the computer and technology. And we were seeing so many kids who'd been pigeonholed already as maybe not having the skill. But as soon as they got into an alternative venue -- they just excelled.... There were so many different avenues for seeing it [children's strengths]. ... When a child produced a product or did a performance, like music, that merited as much consideration as a traditional academic performance.

Or, as Prue said, "Boy! We were seeing kids differently." In this environment the children also came to see themselves differently. Prue commented:

> What we [school staff] were first identifying and recognizing [i.e., children's diverse intelligences], we were clearly articulating to the children. And then the children were saying, 'I can. I can do this.' And, 'I'm good at that.' And 'I can do this.' And then they could also say what their peers could do. So, it was just almost like contagious affirmation. Contagious affirmation.

As these comments reveal, there was widespread belief among the adults and children in the Model Program that the students possessed strengths. Such expectations are often vital to students' success (e.g., Howard & Hammond, 1985). The staff made continual efforts to attend and nurture these diverse strengths through the thematic, hands-on, MI-infused curriculum. In addition, the curriculum incorporated engagement over time and reflection, both necessary to fostering best work. Given all this, it is reasonable to say that the Model Program at Montgomery Knolls encouraged children to do their best work.

<u>Condition 3: Evaluators are Trained to Carry out the Work</u>

To enable teachers to identify the range of youngsters' strengths in the classroom context, and to implement MI-influenced curriculum, teachers were provided with training. In the first year of the grant, Leibowitz reported that teachers had about 10 part-day training sessions. During the second year of the grant there were about five of these. In these training sessions, teachers received information about the theory itself, how it might be applied, and how to use the MI Checklist.

Training to use the checklist was done partly via role playing. Trainers developed little vignettes that teachers would read and act out. The teachers were then asked which of the intelligences they observed in their fellow trainees' role playing, and what was the evidence to support their judgment.

While training was offered, there were some problems associated with its impact and extent. For example, despite training, issues of the "verbal veil" still remained. Jean Barton commented, "...you can train them [teachers]. You can tell them. They can verbalize back. But then when you go look for the application ... [some teachers don't grasp] "what they're seeing." These teachers could not "see" the child's ability, unless the child also had "good verbal ability" and could explain their work to the teacher.

A second problem associated with training was that after the second year of the grant, formal training at Montgomery Knolls was very limited. It was assumed that much of the information had infused the school via the grant council and earlier training sessions. Staff new to the school in the third year and beyond did not have the extended formal training that teachers received at the beginning of the grant.

However, beginning in the third year of the Model Program at Montgomery Knolls, training to adapt the theory into practice and to make the intelligences identifiable in the course of teaching also occurred during day to day work: For example, teachers had access to the district's interrelated art teachers. These teachers helped classroom teachers to program dance and other activities. This enabled the classroom teachers to observe more of their students' strengths. According to one staff member, "It was another piece of training and support to open your eyes and make you think about different ways" to instruct diverse learners.

Through the third year of the grant (late 1992), teachers at Montgomery Knolls could also get training by drawing upon the expertise of two full-time staff members funded by the grant. One helped to devise active, hands-on science units. This sort of curriculum has been advocated as a powerful means of identifying underrepresented youngsters (Leibowitz & Starnes, 1993). The other worked with Hispanic youngsters, while also providing staff support based on her expertise in Montessori, Reggio Emilia, and other early childhood approaches. In addition, she served as a staff-wide resource on multicultural education and curriculum. These individuals provided considerable help to the whole staff, including formal in-services. Their value was tremendous, according to both Hylton and Prue.

During the second Javits grant, there was still ongoing training, though less of it. Hylton took on the title of Javits grant teacher trainer for both Montgomery Knolls and Pine Crest. Because she was also supposed to assist schools throughout the county that were interested in the Model Program, the amount of support she provided for the two Javits-funded schools was limited (MCPS, 1996c, Appendix N).

In sum, ideally, formal training should have been available to staff new to Montgomery Knolls after the first two years of the grant. In the beginning, the training could have also sought to incorporate observations and practice using the MI checklist based on student performances rather than teacher role playing. While training was not sustained at an intensive level, it is reasonable to say that teachers did receive training from a number of sources, and that training was available frequently, if informally, in the school itself. Therefore, the Model Program meets this condition.

Condition 4. Clear Scoring Procedures

Waveline Starnes, who formerly headed Montgomery County's Gifted and Talented Programs, asserted that the county guidelines for identifying children draw on multiple sources of data. Identification is not decided with "a papercutter." As described earlier, there are no absolute score cutoffs on any of the instruments used. However, there are clear county guidelines for Group I students: their scores on three or more indicators fall within the range needed for identification. There are also clear county guidelines for Group III students: they are not identified, because their scores do not fall within or near the range needed on any indicators (MCPS, 1987a).

Some of these indicators, such as the Kough/DeHaan, Renzulli-Smith, and Renzulli-Hartman are teacher checklists. By building in curriculum to address the range of intelligences, teachers had more opportunities to observe behaviors listed on those checklists, such as "Displays unusual talent in music, drawing, rhythm or other art forms" or "Systematically pursues with great absorption one or more special interests..." (Renzulli-Smith, in MCPS, 1987b). As Hylton put it: "See the teacher checklists, that's where the fuzziness comes in. Because the teachers start to see kids differently [with MI]. Then, they started to rate them more highly."

While the teacher checklists may have been where the fuzziness in identification began, it was not the place where it ended. For Group II students, for whom the county guidelines are less clear cut, ambiguity was widespread. The county procedures call on screening committee members to exercise professional judgment in assessing them. Starnes described the committee approach as "... collaborative decision making, which even doctors do. I mean, they do not make decisions all on the basis of what the data says, but make their judgments *using* that data. And I think that's the way you have to select the kids."

For Group II cases, screening members drew more extensively on their experiences with each child. Bulman's comments illustrate how MI became part of the data for decisionmaking in these cases:

> And now, after the grant came and we explored the multiple intelligences, all of that comes into play, and there is a lot of discussion. And if, by chance, the kindergarten teacher or the first grade teacher that has had that same child is on the committee, the test scores are reviewed and then a lot of discussion goes on with the teachers that are familiar with the child. And those areas [multiple intelligences] certainly came into play.

As Leibowitz put it "In Group II we talk about them. And that's where that kind of teacher observation, and work samples, and evidence of problem solving skills come into play." Hylton's comments underscore the equivocal nature of some of these discussions:

> ... supporting data could come forth on a child triggered by an individual teacher's comment or supportive statement, or the opposite: a teacher not supporting a kid might generate support from others. And dogged persistence for an individual could result in [identification], probably like

> a jury room in some way: 'You know I really think we should get back to so-and-so [the debate about a particular student]. I just have this feeling about this kid.' ... 'I really have a feeling about this kid. I've just noticed how -- I don't think we should overlook him.' ... I experienced that happening in both years that I was on those committees. Also on other committees for GT [at other schools].... So, I hear it [advocacy] in other places. You know, the words we've used would be different [at Montgomery Knolls.] So, I can say that. That the MI terms would come up, and supporting information based on experiences with [it].

Hylton's remarks highlight that the basis for advocacy was not detailed or clearly specified. This same approach appears in Leibowitz' remarks on Group II youngsters:

> But what would happen is, a teacher would say, 'look at this project,' 'listen to what the child said when he was solving this particular kind of problem.' ... And we'd go on and describe exactly what the child has done. In that group we would use our professional judgment.

In these cases, intuition reigned. The teachers' intuition may well have been correct, because they did have more experience with the children they were assessing than did DISCOVER or PSA observers. (See Condition 5: Observer Reliability, below.) However, for the purposes of demonstrating clear scoring procedures, intuition is not enough. It is not clear what sort of aids to decisionmaking were used. Though the work brought before the committee was domain-based (see Condition 8), and often involved art work, domain-related criteria or scoring rubrics were not employed. Unlike DISCOVER or PSA observers, evaluators in Montgomery County did not use observer instruments to record or discuss particular pieces of student work. The MI Checklist was based on a range of experiences and work in the classroom, rather than on particular student performances.

While there may have been many other bases upon which decisions were made for Group II students, only two criteria clearly materialized from the data. One of these is "the benefit of the doubt." As Prue noted, a spirit of "contagious affirmation" pervaded Montgomery Knolls. Teachers began seeing strengths in all the children, and this was reflected in the discussions in the screening committee. Hylton commented:

> ...it was, almost, there was a sense they didn't want to let anybody else [go unidentified]. You know, you thought, 'Oh, but I see this and this and this [strength].' I mean it was a very inspiring thing. At the same time, we may have been hugely overidentifying from the point of view of the third grade.... They [Pine Crest staff] didn't look at the kids that way. They didn't see them as an amalgam of their possibilities.

Leibowitz' remarks also illuminate the benefit of the doubt approach to evaluation:

> ...teachers had opened their eyes. And they were seeing kids as gifted in ways that they hadn't looked at kids before. And they were willing to say if they [students] didn't hit the numbers [on county-wide measures] squarely on the head, it was still ok [to identify them]. And that was, I think, a big change from previous years.

Drawing on the benefit of the doubt, whatever strength a child manifested was used as evidence. Given this approach, in some years (1992-1994) about half the students in Montgomery Knolls were identified. "We erred on the side of inclusion," Prue said.

However, not all evidence of strength led to identification. A second criterion, "reality check," constrained the "benefit of the doubt" criterion. The "reality check" was akin to considerations influencing decisionmaking in Charlotte. In both sites, the strengths that teachers found had to be weighed against the real demands that programming for the gifted places on students. Jean Barton said:

> ...many of the teachers [on the screening committee] know that if students are eligible for gifted programming, much of that is going to be very verbal. So that the teachers will sit there and say, 'Well, we see it [a strength], but what are we doing to the child if we put him in a highly verbal [setting]?' ... That's why I'm saying [actual identification] is not a real good criteria of what the teachers are seeing [as strengths]. Because

> they have to continually come back to reality, and say, 'but is this the best learning environment for the child?' And, 'what are we doing to the child?'

The reality check criterion was evident in the case of a teacher who brought to the screening committee a boy's sequence of paintings of trees. The teacher felt the sequence demonstrated an understanding of trees and changes in nature. However, that boy was struggling with language and spent much of his time with a resource teacher. He was not identified because he could not function in a classroom where language demands were high. Similarly, an Hispanic girl who scored in the 99th percentile on the math section of the Test of Cognitive Skills, but who had not yet become functional in English, was also not identified. She was sent to a PADI class for enriched instruction, and was identified a year later.

In sum, while Montgomery County's formal procedures (MCPS, 1987a) were relatively clear, the bases for evaluating students' MI-related strengths were not. Actual criteria used to evaluate the products or processes that students manifest in the course of identification through teaching were not mentioned by anyone who explained the workings of the screening committee to me. Instead, the committee was implicitly guided by two rules of thumb: "the benefit of the doubt" and "reality check."

In order to shore up this aspect of Montgomery County's identification, at least two things could have been done. First, if the staff were going to consider actual products or performances as evidence during the screening committee, they could base decisions on what makes for good student work in various domains. For example, what are the characteristics of work in art, music, science, mechanical construction, or other areas that reveal unusual strength in second graders? The staff at Montgomery Knolls had a leg up on this process, because they had a checklist for the different intelligences, which included behavioral characteristics for each. A next possible step would be for teachers to link products teachers brought to the screening with the behaviors on the checklist. Thus, in discussing a painting, the members of the screening committee could apply characteristics listed under the spatial part of the checklist, e.g., "Shows artistic appreciation, responds to color, line, texture," "constructs and designs visual patterns," or "carefully plans use of space on paper." An event used to highlight a child's strength in the intrapersonal realm could be linked to checklist behaviors like "persistent in self selected activity," "self motivated, independent, and resourceful." These sorts of criteria, available to all those at the table, may have been obscured in an epidemic of "contagious affirmation."

A second possibility was actually to use the existing MI Checklist and make it a formal part of the discussion. Like the Renzulli-Hartman and other checklists, if a child met certain parameters on the MI Checklist, or a combination of parameters on the checklist and other instruments, he or she could then be identified. However, the MI Checklist was never formally a part of the identification, and therefore the bases for drawing on MI in advocating for students remained ambiguous. While the degree of advocacy as Hylton said *is* truly inspiring, it would also be wonderful to see the high identification rates supported by clear criteria. Such criteria were nearly in hand, but not quite grasped.

Condition 5: Observer Reliability

In conjunction with the University of Virginia, Montgomery County's Javits staff did investigate intrarater reliabilities of the MI teacher checklist. A month after teachers at Montgomery Knolls had completed the checklist for all the K-2 students, the teachers were again asked to complete the checklist for 10 randomly selected students from each of their classrooms. Analyses of the 136 randomly selected students yielded "moderately high" intrarater reliabilities for placement purposes (Adams & Callahan, 1994, p. 7). These ranged from .496 in music among first grade teachers to .811 in linguistic intelligence as rated among second grade teachers.

However, digging under the study's statistics reveals problems with its findings. The purpose of an instrument affects how one interprets the scores or analyses resulting from it (Shepard, 1993). Thus, to establish intrarater reliabilities for placement purposes, teachers would have to have completed the checklist with placement decisions in mind. However, teachers used the checklist for a variety of purposes other than identifying giftedness. Leibowitz emphasized that the checklist was designed to provide "examples of the child's [strengths]. It was not: Are you gifted in b-k [bodily-kinesthetic intelligence]? ... It was designed to say ... How do children think? How do children learn? How do they grow?" Similarly, Williams reported that the checklist was used to "observe in terms of children's strengths. Especially [to know] ... what would come about as a result of teaching this particular [curricular] unit, which incorporated the multiple intelligences." Jean Barton said:

> one of the things that we found out was that, in the teachers who really internalized the Gardner model, the checklist was very much functioning as a teacher instructional planning/assessment tool, rather than [only a tool for] identifying -- assessing the various kinds of strengths in kids. Now, I think it did both. But I think it kind of got clouded, and it was intermeshed. I don't think that's bad. I think that really is a good use of the checklist. But I don't think that ever occurred to us when we were first doing it [asking teachers to use the checklist].

If the checklist were used with pedagogical purposes in mind, it might even be reasonable to expect little intrarater reliability. Rather, teachers could be seeing growth and engagement in students, and thus ratings might change from month to month. As Hylton said, "To look at it [the MI checklist] as a static thing and to have it be reliable was, to me, in contradiction to what the whole thing was about in the first place."

However, the checklist to one staffer was "many things to many people." To the teachers, it appeared to be largely a tool for enabling and documenting student change. To Callahan and Adams (1994), who conducted the intrarater reliability study, and to some extent to Waveline Starnes and Jean Barton, the checklist could be used to identify children's strengths. Given these different perspectives on the purpose of this instrument, the intrarater reliability of teachers using the MI Checklist is unknown.

What about the inter-rater reliability of judgments of students' strengths? Would different observers of children in classrooms infused with multiple intelligences tend to draw the same conclusions about a given child's areas of strength? Because this kind of investigation was never undertaken in Montgomery County, the answer is also unknown. In short, there is insufficient evidence to say that the Model Program meets the condition of observer reliability.

Despite an absence of clear criteria to evaluate this range of activities, and though there is no formal evidence supporting reliability, relative to other sites, screening committee members at Montgomery Knolls expressed far fewer doubts about the accuracy of their assessments. This is not self-delusion. Instead, their confidence is based on observations of youngsters over time. In Charlotte and DISCOVER, identification was made by a team whose members were primarily, if not exclusively, from outside the school. Because most team members were not familiar with the students in an ongoing way, team members in Charlotte, and more so in DISCOVER, often expressed doubts about what the children could really do, or whether the work they were scoring represented the students' actual abilities. In contrast, the screening committee of the Model Program was made up of a number of people who really knew the children and how they functioned. As Bulman stated, she had confidence in the committee's decision because of "the fact of what we know about the kids and how they've performed in the classrooms with the teachers who've observed them." As Prue noted, "the grant allowed that whole decision making process to become a much richer discussion about kids, because it brought a lot of data to the setting."

Given this confidence, some measure of interrater reliability might well exist. It might be ascertainable by having those who work regularly with the students in the classroom rate them over time. For example, in most classrooms there was one aide and one teacher. Looking at correlations between these different raters of the same students might be one way to measure the observers' reliability and to provide evidence to support the confidence expressed by screening committee members.

MI-SPECIFIC CONDITIONS

The five general conditions discussed above are needed to associate changes in identification with the assessment procedures. In order to link the assessment with MI, the three MI-specific conditions considered below need to be met.

<u>Condition 6: Assesses Abilities Beyond the Boundaries of Traditional Tests</u>

There is little question that members of the screening committee at Montgomery Knolls considered both traditionally-assessed abilities (linguistic, mathematical, and spatial) and an array of abilities beyond those traditionally measured. As noted in the preceding discussion of observer reliability, teachers reported seeing and developing the diversity of children's strengths within the classroom and then using this information in the screening committee.

The extent to which this information influenced identification for gifted education might be inferred from the increase in students selected between 1992-1994. During this time, the formal measures used by Montgomery County remained the same, but the Model Program's curricular implementation was at its peak (MCPS, 1996a). Thus, teachers had more information about a wider range of students' abilities and used it to identify more students: Bulman's remark quoted earlier highlight this point: "Before the grant actually came and we learned about the multiple intelligences, it [screening] was pretty much based on the test scores of children." Brian Bartels also noted that, before the Javits grant, teachers were "evaluating kids purely -- largely -- in terms of their linguistic and their logical-mathematical intelligences. So they were looking at a very narrow band of intelligence. They are looking at the child much more holistically now." Given this, the Model Program meets the condition of assessing abilities beyond those traditionally tested.

Condition 7: Intelligence-Fair

An intelligence-fair assessment allows children to demonstrate their abilities in media pertinent to the problem solving at hand. Thus, an intelligence-fair assessment of musical ability might entail playing musical instruments or singing rather than writing or talking about how a song sounds. (See Chapter 1.)

In general, the Model Program at Montgomery Knolls did assess students in intelligence-fair ways. Bulman's remarks underscore the intelligence-fair nature of identification through teaching:

> I guess we're just so much more aware of how the children work best, because we've offered them opportunities now to show us.... And now that we've set up these activities, and you really see how they do their best, that's something that's more in the forefront of our minds now than it ever was. And we want to know: Show us how. Let us see it.

While there is clear evidence from Bulman and others that intelligence-fair practices were in place, there is some evidence, not surprisingly, that the practice was uneven. Hylton reported:

> [T]here wasn't one set of [assessment] activities. Which means that it [identification through teaching] was in teachers' heads. And those teachers who got it [i.e., offered MI-infused practices], got it [i.e., saw children's diverse strengths], and those who didn't, didn't. And what does that really mean? If you didn't set up the opportunities, then you may not see it [the children's strengths].... I think it differed dramatically from a Karen [Bulman] and some of the others [who worked with] very little ones, where there was decreasing print and more active observing that was required, to the second grade, where print wasn't the only thing, but you used it.

As Hylton's comment indicates, linguistic skills still played a large role in some classrooms. Jean Barton noted that "the chief thing that we have struggled with is the verbal halo effect, if that's what you want to call it." Despite training, Barton felt teachers still believed that "If they [children], can't talk about it, they don't know it."

Linguistic capabilities not only influenced teachers' perceptions of children's strengths in the classroom, they also entered into the actual decision making during the screening committee. As noted in the discussion of scoring procedures, the "reality check" rule of thumb essentially coupled strength in mathematics, the spatial realm, or other areas, with a degree of English language competence needed to function in demanding programs. Students' ability to communicate in English influenced teachers' placement decisions. This may have undermined increases in Hispanic students' identification. Hispanic youngsters were the only group in the Model Program to be consistently underidentified. (See Table 4.1.)

On the other hand, in Montgomery County, youngsters could be identified without strong second-order skills, via teacher ratings on the Renzulli-Hartman scale, the Kough/DeHaan teacher behavioral checklist, and teacher advocacy. Most teachers became willing to advocate for students' strengths as represented in a range of media from movement to music to paint. As a result, more youngsters were actually identified.

In essence, with regard to being intelligence-fair, Montgomery Knolls' approach falls between that of DISCOVER, in which children can be identified without demonstrating competence in language or notation (via Pablo® and tangrams) and Charlotte, where second-order notational skills were essential for identification. While Montgomery's approach falls short of a theoretical ideal, there was still a reasonable possibility for youngsters to be identified without second-order skills and with adequate, rather than exceptional, language skills. Thus, it is reasonable to credit the Model Program with meeting the condition of being intelligence-fair.

Condition 8: Domain-Based

According to Gardner (1983), intelligences are recognizable only in the context of cultural practices or "domains." Thus, to evaluate whether a youngster has unusual bodily-kinesthetic abilities, it is necessary to see those abilities as they are employed in sports, dance, model building, or other domains that draw on large and/or small motor skills. For the most part, the Model Program at Montgomery County did evaluate children's strengths in the context of domain-based activities.

A number of Model Program strands enabled a focus on domains. For example, as noted in the description of program elements, learning centers drew on domains like drama, music, and movement.

The program also focused on thematic units, which integrated learning from a variety of disciplines/domains in order to foster understanding. Thus, in the cowboy unit, children learned cowboy songs, studied brands used by ranchers, and designed and drew their own brands. In a unit on birds, children got regular opportunities to be birders and develop the skills of a proto-biologist: observing with binoculars, drawing what they saw, and developing graphs based on the frequency of their observations.

Complementing the integration of disciplines was the Model Program's emphasis on "active learning." Real domains are actually practiced -- not just acquired by reading and writing. Active learning fostered real domain practices: observing birds, recording one's observations in drawing, and constructing one's own graphs,

Domain-related work was enabled by staffing. As noted earlier, one of the grant-funded teachers had expertise in developing science curriculum. She helped the whole staff to develop domain-relevant science activities. In both schools there were teachers of art and music on almost a full-time basis. These disciplines are rarely transmitted primarily via words and paper; instead, they typically rely on domain-related activities and materials.

Because of thematic units, active learning, and staffing, Montgomery Knolls' teachers were able to bring to their advocacy at the screening committee evidence of students' strengths as demonstrated within domains. They could also draw on domain-related performances in completing the Kough/DeHaan checklist and the Renzulli checklists. Thus, though not all the instruments used to identify children at the screening committee were domain related (e.g., the Test of Cognitive Skills, the Raven's), it was still possible to identify youngsters on the basis of domain-based performances. Therefore, it is reasonable to credit the Model Program with meeting this condition.

CONCLUSION

The analysis of the Model Program against the eight conditions reveals that Montgomery Knolls is the only site in this study to meet all three conditions pertaining to MI: Identification did draw upon more than the three traditionally assessed intelligences; it was intelligence-fair, and it was domain-based. Thus, it is reasonable to link the Model Program's identification effort to MI. However, the analysis of the Model Program against the five general conditions indicates that there were neither clear scoring procedures, nor observer reliability. Thus, though the Model Program's assessments can be linked to MI, it is not reasonable to draw inferences from these MI-influenced assessments to increases achieved by the Model Program.

Despite this, the Model Program had a number of strengths. First, it relied on a much greater and more representative sample of students' work than the other sites. This likely yielded a more veridical picture of students' abilities. Second, the Model Program brought into the identification process classroom teachers who knew the children well. Third, the great majority of the staff came to see that all children had strengths. Therefore, more children were provided with increased access to advanced programming.

One difference between the Model Program and the two other sites is that the program did not meet its aim of increasing the identification of underrepresented youngsters. Why was there basically no difference in equity in the one site that actually implemented MI, while the two sites that did not implement MI achieve greater equity?

The answer, I believe, has less to do with the content and methods of the Model Program than with situating it at Montgomery Knolls and in Montgomery County:

One limitation of this setting was the extent to which MI's role in identification was limited by county guidelines. As discussed earlier, Group I and Group III youngsters were identifiable (or not) on the bases of screening instruments already used throughout the county. Thus, the decisionmaking for only one-third of the youngsters -- those in Group II -- was heavily influenced by the new MI-influenced approach.

A second limitation on equity was the influence of Montgomery County's existing curriculum. The need for students to function in programs with high language demands likely dampened the identification of Hispanic youngsters, who were consistently underidentified before and after the grant. (See Table 4.1.)

Third, with regard to African American students, Montgomery Knolls had already made strides partly via the PADI program. In 1989, the year *before* the Javits grant was awarded, the identification rate for African American students matched their presence in the wider school population. In essence, there was a ceiling effect for African American students. MI was introduced in a school where a large additional benefit for these youngsters could not be demonstrated. As Prue said, "we were really doing a satisfactory job prior to the Javits coming in and making these changes." Donnelly Gregory, the PADI coordinator, noted Starnes' determination to achieve "demonstrable significant results of this project. The unfortunate thing is that she never had any way of designing a way to look at it." Similarly, in a separate conversation, another staff member said, that with the grant councils and think tanks, the designers "expended energy in organizing and planning it [i.e., the Model Program]. Then they tried to force data to fit" that effort.

Though both the staff members above were critical, they also spoke of the potential of the program for positive change. As one remarked, "it was a good project, but it could've been a great project." For at least a few years, more youngsters overall were identified. In addition, in the everyday workings of the school, something very powerful happened: Many teachers were finding new ways of seeing and developing youngsters' strengths. Children came to see themselves as able in many different areas.

As Jean Barton observed, "I think we did assess giftedness. I think we got some people to broaden their view of what giftedness was. The problem was that we didn't impact the whole system ... the whole county." Though it was designed in part to serve as a model for good practice for the county, state, and neighboring metropolitan area (MCPS, 1989), the work at Montgomery Knolls exercised little systemic influence on identification anywhere: The MI Checklist, a concrete and compact instrument, and an obvious candidate for formally supplementing existing screening measures in the county, was never more than loosely tethered to the identification process. In contrast to Charlotte's yellow card, or DISCOVER's Observer Notes, the behaviors on the MI Checklist were not formally used, even at Montgomery Knolls. The adoption of the checklist and other elements of the Model Program at Pine Crest was described by Hylton as "sporadic," despite Javits funding to foster MI-influenced approaches there. Moreover, since the departure of Pam Prue from Montgomery Knolls the impact of MI even there has diminished. Leibowitz reported hearing from several sources that some teachers new to Montgomery Knolls do not even know what MI is. If so, it is hard to imagine how they might be using the theory in identification.

In essence, the Model Program made a difference, but its impact was not very wide and was not sustained. The MI-influenced Model Program was like a thoughtful visitor to Montgomery Knolls. For a while, it fostered some different activities and conversation. But when the visitor left, that conversation's echo diminished.

What were the reasons that Montgomery's Model Program never became a model? Why was the work confined to a single school, even though the county's proposal to the Javits Program stated dissemination as an important aim of the work? What are the possibilities that within Montgomery Knolls the general conditions that were not met might yet be? The contextual issues that took Montgomery Knolls and the other sites to their current, respective circumstances are explored in the final chapter.

1. Starnes believes that the disproportionately low number of minority students in these programs is partly due to discomfort among some African American students about leaving their home schools for a competitive magnet environment, as well as by efforts of the students' home schools to retain talented students.

2. Though Asian students in the district are largely from more well-to-do circumstances (Eaton, 1996), staff who worked in Montgomery Knolls reported that that school's Asian population tended to be recent immigrants, without much in the way of economic resources, and often weak in English language skills.

3. These figures are from the final report submitted to the Javits Program. They vary markedly from figures in earlier documents. According to Starnes and Leibowitz (1993, p. 32) "The number of second graders formally identified as gifted and talented using the standard county-wide multiple-criteria grew from 17 percent the spring [1989] before the grant was in place to 23 percent [1990], and then 42 percent for each of the remaining two years [1991, 1992] of the first grant."

4. Follow-up calls to Pam Prue and Brian Bartels, a psychologist at Montgomery Knolls who helped analyze the data, have not yielded a strong explanation for the large drop in identification of white students that occurred in 1994. Neither of them is quite sure why this happened, but both speculated that some change in student school assignments may have taken place that year.

5. Despite being assured by the acting director of Enriched and Innovative Instruction that I would be able to observe screening committees in action, over 20 requests made between January and April, 1996 to arrange for this were never granted. Because of this, information pertaining to the decisionmaking process is based on interviews with people who participated in the screening committees and on documentary data.

Chapter 5
REVOLUTIONARY ASSUMPTIONS, EVOLUTIONARY CHANGE

INTRODUCTION

In many ways, the work that was described in each of the three preceding chapters began with revolutionary assumptions. Each of the sites overturned the common operating assumption that poor and minority youngsters will be disproportionately underrepresented in programs for the gifted. In addition, Charlotte's and DISCOVER's designers set aside their home states' tradition of identifying gifted youngsters on the basis of standardized testing. In all the sites, new identification methods were inspired by Gardner's theory of multiple intelligences. According to this theory, intelligence entails solving problems or creating products valued in a culture, and traditional psychometric tests measure only a limited range of intellectual strengths (Gardner, 1983).

In Montgomery's Model Program, an assessment process incorporating such assumptions was accompanied by increases in the number of youngsters identified. For Charlotte and DISCOVER, there have been increases in the identification rates of traditionally underserved students.

Drawing on a framework of eight conditions, the preceding three chapters analyzed whether it is reasonable to associate these outcomes with each site's identification methods and with MI. The first five "general" conditions are needed to make inferences about individuals' abilities from any assessment. These must be in place to associate claims about improved rates of identification with the assessment itself. The second three conditions are "MI-specific." These are needed to link the assessment to MI theory. (See Chapter 1.) Table 5.1 below summarizes the analyses of the three sites with regard to these eight conditions. (This table is not intended to create a scale; it only outlines the analyses detailed in the preceding three chapters.)

Table 5.1: General and MI-Specific Conditions Met by the Three Sites
(A summary of the detailed analyses for each of the sites)

	DISCOVER	Charlotte	Montgomery Knolls
General Conditions			
Condition 1 Children Understand Tasks	yes	yes	there are no tasks; understanding of curriculum varies
Condition 2 Children are Encouraged to do their Best Work	yes	yes	yes
Condition 3 Evaluators are Trained to carry out the work	-	yes	yes
Condition 4 Clear Scoring Procedures	-	-	-
Condition 5 Observer Reliability	-	-	-
MI-Specific Conditions			
Condition 6 Assesses Abilities Beyond Traditional Tests	-	-	yes
Condition 7 Intelligence-Fair	yes	-	yes
Condition 8 Domain-Based	-	-	yes

As Table 5.1 illustrates, each site met some, but not all of the general conditions. Without having met all the general conditions, it is not yet possible to associate improved rates of identification with the assessment procedures. That each site met only some of the general conditions is not surprising: At the time I visited the sites, they were in operation between three and six years. Developing tasks, instruments, scoring methods, and training procedures -- while simultaneously implementing these in schools and investigating them -- is complex and time consuming. Given this, Charlotte-Mecklenburg's progress is impressive: As discussed in Chapter 3, the Problem Solving Assessment is on its way to achieving clear scoring procedures and observer reliability, and thus meeting all the general conditions.

More surprising is that all three sites explicitly sought to build identification on MI and yet none was actually drawing deeply on the theory. Charlotte meets none of the MI-specific conditions. DISCOVER employs one: intelligence-fair approaches. Montgomery Knolls' actually meets all three MI-specific conditions. However, there MI never became a formal part of the school's identification procedure. (See Chapter 4.)

Why did the implementation of MI in the sites take shape the way that it did? One possibility suggested by Gardner (personal communications, April 1996, April 1997) is that the designers did "not really understand my theory." Evidence supporting this hypothesis is limited. Though none of the sites spoke directly with Gardner about the assessments they were formulating, designers in Charlotte and Montgomery County did consult with Project Zero staff members who developed the Spectrum activities. (See Chapter 1.) In addition, designers in Charlotte and Arizona have undertaken widespread readings of Gardner's books and articles. For example, as discussed in Chapter 2, Maker and Nielson understood and drew on Gardner's notions of first- and second-order knowledge. They felt (as did the designers in Montgomery County and Charlotte) that "we can identify different strengths than have been traditionally identified." That

Charlotte and DISCOVER have difficulty assessing more than the three traditional intelligences is not explained by a lack of understanding of the diverse intelligences Gardner posited in MI theory.

A similar situation holds with respect to being intelligence-fair. As Table 5.1 indicates, Montgomery Knolls and DISCOVER meet this condition. The designers in Charlotte certainly understand the importance of intelligence-fair assessment. As noted in Chapter 3, the designers said they tried to use hands-on materials and "to get as far away as we felt comfortable from paper and pencil." Thus, a lack of understanding does not explain the designers' difficulty in making the PSA intelligence-fair.

The only area of misunderstanding, as I will highlight in the discussion of DISCOVER below, concerns the notion of "domain." Nevertheless, overall, it is not a fundamental misunderstanding of MI that has limited the implementation of the theory in the sites to this point. The shape that MI -- and other theories and ideas -- takes in educational practice does not depend only on an understanding of the theory. Even a robust understanding may not result in a robust implementation. (This situation is illustrated below in the case of Montgomery County). An understanding of the theory is necessary but certainly insufficient. In many crucial ways, as I will detail in this chapter, adapting the theory depends on features of the context into which the theory is being fitted.

The findings from the three sites, summarized in Table 5.1, provoke many questions. Among those addressed in this chapter: Will the designers be able to incorporate MI theory more firmly? Will they be able to meet the general conditions, so that it is reasonable to make inferences about students from their assessments and to associate changed outcomes with the assessments? In essence, what are the possibilities for strengthening these assessments, so that the strengths of more underrepresented youngsters can be recognized, developed, and justified on methodological grounds?

To examine these questions, I highlight forces that shaped and tempered the revolutionary assumptions underlying the assessments in each site. In a more speculative vein, I consider how these assessments may evolve over time given the contexts in which they operate. Finally, given the forces acting on these assessments, I propose steps that policymakers may take to support the development of more equitable approaches to identification.

DISCOVER III

As detailed in Chapter 2, DISCOVER has been able to meet two of the five general conditions: Children do understand the assessment tasks, and children are encouraged to do their best work in the assessment. Other aspects of DISCOVER diverge from the general conditions. Specifically, the assessment does not rely on adequately trained observers, and the scoring procedures are not clear. Given these circumstances, observers' reliability has not yet been established, despite suggestions to the contrary (e.g., Giffiths, n.d.; Nielson, personal communication, February 18, 1997).

As for the three MI-specific conditions, DISCOVER does employ intelligence-fair approaches. That is, children can be identified without having to translate their abilities primarily into notations or language. However, the assessment does not extend beyond language, mathematics, and spatial abilities, the three areas traditionally tested by standardized measures. It also does not meet the condition of being domain-based: The majority of tasks, including tangrams, Pablo, and the math worksheet are not grounded in cultural practices. To understand why DISCOVER meets some conditions and not others, it is helpful to consider how it was shaped by the context in which it developed.

State Policy

Though DISCOVER assessments are now used in several states as well as Canada, the assessment's Arizona origins have left their imprint. According to Arizona state policy, students must be identified and provided with services if they score at or above the 97th percentile on one or more state-approved, nationally-normed standardized tests of linguistic, quantitative, or non-verbal (typically spatial) abilities.[1] This state policy appears to have constrained DISCOVER from going beyond the three traditionally tested abilities. As Aleene Nielson put it, "Because the state would recognize excellence in those three areas, those were the three areas that June wanted to include in the assessment" (personal communication, February 18, 1997).

It is also possible that the state's funding policy may have limited the range of assessed abilities. Though there is no limit on the number of students that a district can identify, the state provides funding to serve only up to three percent of a district's students (Arizona Department of Education, 1992). If DISCOVER assessed the full range of intelligences, it may well have identified significantly more youngsters than state and local finances could serve. Given the limited financial incentives coming from the state, districts would be less likely to adopt an assessment that might dramatically expand the number of students who are identified.

History

History has influenced DISCOVER in at least two ways. First, some of the materials used in DISCOVER stem from earlier efforts and collaborations. As noted in Chapter 2, tangrams were first used some months before DISCOVER was launched as part of Maker's and Rogers' work to enhance the identification of Hispanic youngsters in the Tucson Unified School District. The Pablo® task drew on materials Maker had received some years earlier from an educational products and services company, which she had already applied in other educational settings. The math questions, especially those nearer the closed end of the problem-solving continuum, are inherited from the sorts of paper and pencil problems schools traditionally formulate.

Both Pablo® and tangrams lend certain strengths to the assessment: they are hands-on tasks that help make the assessment intelligence-fair. They are also interesting materials that engage most of the youngsters and encourage them to do their best work. However, assessing the products resulting from this engagement is difficult from the perspective of MI theory. Because they are not domain-based, it is not clear how they should be judged. Partly because of this, and because there is only a classroom-based reference group, the scoring remains unclear.

Alongside the history of DISCOVER itself, the history of psychometric assessment influences DISCOVER. Maker, and other assessment designers, find that domain-free tasks can enable youngsters who may have had few encounters with a given domain to demonstrate their strengths, without suffering in comparison to youngsters with richer experiences. This effort to control for differential experiences by using novel tasks is fundamental to traditional intelligence testing. These practices also lend DISCOVER some strength: they do not wholly sever the assessment from the psychometric mainstream. However, the de-emphasis of culturally valued problem solving does weaken the tie between DISCOVER and the MI.

Curriculum-Assessment Link

Another constraint limiting DISCOVER to the traditionally measured abilities is the current curriculum in most schools. From the perspective of MI theory, all intelligences are equal (Gardner, 1983): One should not be labelled intelligent -- or, by extension, gifted -- on the basis of strong performances using some of the intelligences but not others. However, in school the 3Rs remain central as either subject areas or the means by which subject areas are presented. The 3Rs are linked to most of the formal assessments that occur in school. (The extracurriculum and "specials," which can more readily engage abilities beyond linguistic and logical-mathematical, are often not formally assessed.)

Even though Maker initially planned to devise assessments for each of the intelligences, schools' existing curriculum drove DISCOVER in the direction of traditionally-measured areas. As noted in Chapter 2, Maker explained that to assess a fuller range of intelligences:

> First of all, you have to get people to believe that musical and bodily-kinesthetic [abilities] would be important to assess, because they [most educators] don't see their task as having anything to do with development of bodily-kinesthetic and musical intelligence.... And so, if you're going to develop an assessment, you start where you think somebody's going to use it. That's my attitude. Start where you think somebody's going to use it and then expand.

Maker's comment also suggests another link between the curriculum and the development of DISCOVER. Her assessment needed to identify youngsters who could be served; if children were identified on the wide range of intelligences, there would be a mismatch between some of the identified students and the curriculum of the classrooms into which they'd be placed. Nielson's comments touch on this point as well: "if we place children on the basis of alternative assessments and then we continue to offer the same kind of program, their strengths are going to be ignored." By starting out in the typically assessed areas, DISCOVER identified youngsters who stood a reasonable chance of having their needs met.

Resources

Among the obstacles to broadening the assessment beyond the traditional three abilities are human and financial resources. DISCOVER received $796,548 between mid-1992 and mid-1996 to work in 9 different LEAs across Arizona (Barnes, personal communications, 1997). This was not a great deal of money, given the scope of the work: As detailed in Chapter 2, DISCOVER assessments of spatial, linguistic, and logical-mathematical abilities are labor intensive. It takes four or five people an entire day to administer the Pablo®, tangrams, and storytelling and to reach decisions about the performances of a single classroom of children. Funding personnel to go to the sites and carry out these assessments may not leave additional resources for the development of tasks beyond the three traditionally tested areas. Furthermore, DISCOVER resources had to be spread among the assessment activities, staff training, and curriculum development.

Along with funding limitations for developing the assessment, DISCOVER is constrained by the finances of schools and districts that may want to administer it. Maker states that DISCOVER is reasonable from the perspective of cost: It runs somewhere between a program of individual testing and standardized group testing. However, adding other tasks (without also revising the existing instruments and scoring procedures) would increase time and labor costs, tipping the balance away from financial viability. Clearly, the higher the cost, the less attractive DISCOVER will be to schools and districts.

University/Research Context

DISCOVER's developers, including Maker, Nielson, and Rogers are all steeped in the actual practice of their assessment in schools and are knowledgeable about the schools in which the assessment is being implemented. Still, unlike their counterparts in Charlotte and Montgomery County, they are denizens of a research university and not a school system. As a result, Maker and her colleagues are far less likely to be criticized by, or need to respond to, teachers, district administrators, citizens groups, or parents. In essence, the designers' university setting helps to insulate them from a range of potentially useful critique.

Alongside such insulation, university-oriented research aims may foster resistance to substantial revision of the DISCOVER assessments. In particular, a great deal of time and energy have been spent collecting data based upon the tasks, procedures, and instruments. There is, therefore, a cost in seriously altering the identification process: Such revisions disrupt the possibility of longitudinal studies and make statistical analyses more complex. For instance, when asked why the story-telling task was not moved from just before lunch to a less distractible time, Rogers said she believed the tasks order was maintained to prevent problems with data analysis.

An additional problem is that the research conducted by DISCOVER is not geared to inform revision or modification of the instruments. Rather, it seeks to "validate an innovative procedure for identifying gifted minority students..." (U.S. Department of

Education, 1994, p. 6). Further, these validation studies are largely conducted by Maker's graduate students using data already collected via the tasks, procedures, and instruments that Maker and her colleagues have devised. By directing the research primarily at validation and having it conducted by her own students, Maker receives limited feedback about the tasks, procedures, and instruments themselves.

To illustrate, Griffiths' (n.d.) reliability studies are not weighed against the actual practice of the assessment in the field. (See Chapter 2.) Griffiths instead suggests observer reliability exists, even though inadequate observer experience is widespread and there are deficits in training as well. The validation agenda blindsides the designers: In her response to the issues of training and reliability presented in Chapter 2, Nielson (personal communication, February 18, 1997) continued to argue that "Overall, inter-rater reliability is very high as Sarah Griffiths has shown...."

In short, DISCOVER's university/research context has screened out potentially useful sources of information. Lacking the degree of feedback and scrutiny of their Charlotte counterparts (see below), the DISCOVER designers have not had to shore up their training or simplify their instruments. Their dedication to validation may be steering them away from modifying their existing approaches in ways that could make for clear scoring procedures (e.g., by eliminating unnecessary behaviors from their checklists). Their validation effort has also persuaded them that the assessment is already reliable, when this is not yet a reality.

Leadership

DISCOVER is very much shaped by the visionary and pioneering leadership of C. June Maker. As noted in Chapter 2, Maker has a long history working to identify gifts and talents in people typically overlooked by traditional assessments. She has succeeded with the DISCOVER assessment in identifying strengths in youngsters who have been unnoticed in the past. This success is testimony to her conviction that strong problem solving skills exist across cultures, races, and classes.

As powerful as this conviction is -- and it is one that has motivated virtually all of my own work for the last decade -- there is yet a need to tether assessments stemming from it to clearly defensible methods. Such methods will support the work's moral foundation: defensible methods can enable DISCOVER to withstand the scrutiny of critics who are not predisposed toward either alternative assessments or equity in gifted education. DISCOVER then might accomplish even greater equity.

Maker's perspective on DISCOVER and mine are not aligned (personal communication, February, 1997). In particular, she believes that scoring is clear and that my analysis reflects the idiosyncratic approach used by the team I observed in Chinle. She thought that if I had observed her instead, my findings would be quite different. This is certainly possible.

On the other hand, DISCOVER assessments are not primarily administered by Maker. Rogers commonly led the assessment teams. She and the other team members I observed worked extremely hard and thoughtfully throughout workdays that lasted 10 and more hours. If a group led by a highly experienced observer and Maker colleague is not performing in line with Maker's vision, it is reasonable that educators elsewhere who have adopted the DISCOVER process are also not conducting the assessment in a way Maker would find satisfactory. Their performance is impeded by the complex and time-bound nature of their assignment. The findings presented here about scoring procedures and reliability speak to DISCOVER's design and implementation at least as much as it does about any particular team.

In short, while Maker's visionary leadership has enabled new ground to be broken, to make this terrain more widely traversable, this vision must be informed by existing challenges: the need for more training and clearer scoring procedures; the recognition that, without these, observer reliability will be hard to achieve.

LOOKING INTO THE FUTURE OF DISCOVER

Given the forces that have shaped the current form of DISCOVER, what are the possibilities that this assessment will meet the general and specific conditions it does not yet meet? With regard to expanding beyond the three traditionally tested areas, DISCOVER is constrained by state policy, curricular traditions, and limited resources to develop and implement assessments for different intelligences. Given this, I believe it will be difficult for DISCOVER to expand and implement assessments beyond the traditional three areas in the near future.[2]

However, this expansion may not be the best use of DISCOVER's resources. Not assessing other areas will continue to place the work at odds with Gardner's theory. However, even without expanding the assessments, DISCOVER has enabled teachers to see children in new ways. Nielson reports:

> I think one of the really exciting things as we get the assessment out there, and teachers begin to see things that students can do that they didn't believe students can do, they're getting a much better picture of the

problem solving strengths of the children they work with. And it gives them a different perspective.

Similarly, Maker reported that a fourth grade teacher told the DISCOVER team when they entered his classroom, "'There aren't any gifted kids in there.'" But after watching the children's performances on tangrams, "he was surprised that his children had done so well, and he thought he should take a look at some of those children who had done them."

Though children may not be identified on as broad a basis as MI would support, by meeting the intelligence-fair condition, youngsters who might otherwise go undetected get noticed. As teachers in classrooms begin to see children differently and come to appreciate that children may have an array of strengths, they become more likely to provide opportunities that would engage and develop these youngsters' abilities.

Thus, from a pragmatic standpoint, it may not be necessary for DISCOVER to assess other intelligences. Their intelligence-fair tasks are enabling more youngsters to gain access to the kinds of challenging curriculum that was typically denied them.

As for meeting the condition of domain-based assessments, there are several obstacles that make it unlikely. First, Maker feels novel tasks do not disadvantage youngsters who may not have had exposure to rich learning environments. This is a position that many other assessment designers have taken; novel tasks have a long history within the psychometric mainstream.

Another obstacle to establishing domain-based assessments is misunderstanding about what a domain is. Gardner (1993b) noted the term was not clearly defined when MI was first posited. However, prior to the start of DISCOVER III, the meaning of a domain was clarified as a "discipline or craft that is practiced in a society" (Gardner, 1991a). The idea of assessing intelligences via performances within culturally valued domains was highlighted more than a decade ago (Walters & Gardner, 1986). Instead of trying to evaluate behaviors related to these practices, the DISCOVER checklist is said to build upon the "core capabilities" for each intelligence (Nielson, personal communication, February 18, 1997). Unfortunately, core capabilities are not readily observable in school settings: they are basic neural mechanisms that are "triggered" into processing information by particular kinds of stimuli (Gardner, 1983, p. 64). (For example, spoken sentences are automatically processed into discrete words by individuals functioning in their native languages; this is not an observable behavior).

Modifying DISCOVER to incorporate domain-based practices is also complicated by the commitment to the assessment as it is now configured. For example, the checklist behaviors are derived from extensive observations of how diverse youngsters solve largely novel tasks, rather than domain-based tasks. Adopting more domain-based assessments would require changing the checklists through which much data have already been collected. Should the DISCOVER designers decide to alter their course toward domain-based assessments, this would help them to develop meaningful, culturally-sensitive tasks as well as criteria around which to judge the tasks.

The three MI-specific conditions are ways of understanding whether it is reasonable for DISCOVER to be associated with MI theory. While it would be helpful if DISCOVER could adopt more domain-based tasks, this and the other MI-specific conditions are not essential to making inferences about students' abilities from the assessment. Such inferences rest on meeting the five general conditions. If these were met, Maker would be able to associate changes in outcomes with the DISCOVER assessment, even if the assessment were not closely linked with MI.

With regard to observer training, what is still needed is for all observers to get the training required for their demanding role. Especially if Maker presses for it, DISCOVER could ensure all observers are well-trained and have solid experience before going out to conduct assessments. As suggested in Chapter 2, by drawing on an apprenticeship model, Maker could both ease the burden now placed on observers while giving novices greater training.

Reliability might also be achieved, if training was made mandatory and there were greater efforts to use experienced observers. With these features in place, the high inter-observer reliabilities found by Griffiths (n.d.) among experienced observers should be achievable for the observer team as a whole. (See Chapter 2.)

A condition DISCOVER is less likely to meet is that of clear scoring procedures. For this to occur, the designers would need to believe there is work to be done in this area. Maker and Nielson (personal communications, February 1997) already feel scoring is clear. This argument is partly based on recent findings by Catherine Seraphim, a Maker doctoral student. According to Nielson, Seraphim has found that identified students accrue more checkmarks on the DISCOVER checklist than those who are not identified. While this provides some post-hoc statistical support, the actual scoring procedure for the observers remains quite unwieldy: Data tables from the dissertation sent by Nielson show that about a third of the 90+ checklist items are rarely if ever used. As discussed in Chapter 2, the observers are not always sure what the checklist items mean and many items are redundant. Certainly, there is room to simplify this scoring procedure.

A further obstacle to achieving clear scoring procedures is DISCOVER's commitment to classroom-based reference groups. Such an approach has strengths: it does not compare children in one classroom to those elsewhere who may have had quite different and possibly richer experiences. Thus, it likely allows more children from underserved populations to be identified. However, this approach has costs: it is hard to anchor observer judgments against a reference group or set of performance criteria. It is also hard to understand what any one child's performance actually says about the child's abilities. There is room for a middle ground: DISCOVER could develop norms that were still local, but that were above the classroom level. For example, a norm might be established for youngsters in the Chinle Unified School District. (See Chapter 2.)

Whether or not efforts are undertaken to revise DISCOVER, important outcomes have been realized from this assessment. More children have been identified and served, even if the basis for this was not as clear as it could be. It is worthwhile remembering that traditionally used methods of identification, despite their reliability and other technical merits, are also not wholly satisfying: their predictive validity for actual adult accomplishment is far too low to continue to justify denying access to challenging curriculum to youngsters, especially since these lost opportunities contribute to existing racial, ethnic, and class inequities.

Another positive outcome is that DISCOVER's intelligence-fair approaches, coupled with its professional development efforts, have enabled many teachers to see strengths and potential in their students. This has led some teachers to develop their practice and thereby to enhance children's curriculum and classroom experiences. Examples noted above by Maker, Nielson, and educators at the two Arizona schools I visited highlight this point.

A key contribution of DISCOVER is that its pioneering methods cleared new ground upon which efforts like Charlotte's PSA are built. According to Nielson (personal communication, February 18, 1997), the DISCOVER manual will soon be published. This should also spur other fruitful adaptations.

CHARLOTTE-MECKLENBURG'S PROBLEM SOLVING ASSESSMENT

Though Charlotte-Mecklenburg's Problem Solving Assessment was based on the DISCOVER III model, it has evolved into something quite different. With regard to the MI-specific conditions, the forces that have shaped the PSA have taken it farther afield of MI theory. Like DISCOVER, the PSA does not extend beyond the three traditionally tested areas. However, unlike DISCOVER, the PSA does not meet the condition of being intelligence-fair. The assessment does employ some intelligence-fair tasks (e.g., storytelling, Pablo®, and tangrams). However, these constitute neither the majority of the tasks, nor is it possible for students to be identified on the basis of their performances on the intelligence-fair tasks. (See Chapter 3.) The same situation holds with regard to being domain-based: Some of the PSA tasks are domain-based (e.g., storytelling, storywriting, and the map). However, children cannot be identified on the basis of domain-based tasks.

As for the general conditions, the PSA does use well-trained observers. Because studies of observer reliability have not yet been conducted, it is not possible to say the

PSA has achieved observer reliability. However, given observers' training and high level of experience, observer reliability might well be demonstrated once studies are undertaken. At the time of my visit, clear scoring procedures were not in place. Since that time, scoring rubrics have gone into development. If this work continues, the PSA may realize clear scoring procedures in the next year or so. Given this, the PSA may well meet all the general conditions required to make inferences about students from the assessment itself. If so, Charlotte-Mecklenburg will be able to justify enhanced equity on the basis of the PSA, even if these gains are not much related to MI theory. To understand the likelihood of meeting the conditions that are not yet met, it is useful to understand the forces that have shaped the PSA to this point.

State Policy

At the time the PSA was being developed, state policy in North Carolina, like that in Arizona, reinforced the assessment of logical-mathematical, linguistic, and spatial abilities: According to North Carolina policy children were officially identified as gifted primarily via standardized achievement tests and IQ tests, with some additional points for school grades. (See Chapter 3.)

Though the PSA has become Charlotte's primary means of identifying youngsters for gifted education services, parents can still request the state's testing procedure. According to Udall and Reid, allowing the state method of identification puts Charlotte in compliance with state law. Being in compliance enables the district to receive additional funding for 3.9 percent of the total student population, which is applied to the Program for the Gifted budget (see "Resources," below).

Given the state method of identification, Charlotte's gifted programs were populated by students with notational abilities and strengths in the three traditionally tested areas at the time the PSA began. Teachers of the gifted were used to working with these students, and this put some pressure on the designers to continue selecting youngsters with similar abilities. This helped to keep the assessment restricted to linguistic, logical-mathematical, and spatial abilities. It also encouraged the addition of notation-oriented tasks, which made the assessment less intelligence-fair. (See "Curriculum-Assessment" Link, below.)

<u>History</u>

The clearest historical influence on the PSA is the work of DISCOVER. When members of Charlotte's task force began exploring other efforts to apply MI to identify underserved youngsters for gifted services, they looked largely to Maker.

In comparison to Montgomery County's approach, Maker's work was attractive because of its well-articulated assessment tasks. As Reid said, the members of the gifted education task force felt they "needed something that was a practical, observable procedure with children in the intelligences." Similarly, Udall commented that "the problem in using a theory like this one is that it's not built for school systems. It's not built for all this stuff around the pragmatic aspects of identification. I mean, that's why we use IQ tests." She noted that Maker's "trailblazing really helped us ... in terms of how do you take this theory and turn it into practice." Thus, Charlotte began assessing in the same three traditional areas of ability as DISCOVER and still uses some of the same tasks. (See Chapter 3.)

Alongside the history of the assessment tasks, there is, as Udall noted, the larger history of Charlotte-Mecklenburg and its efforts at desegregation. The district is a place where the allocation of educational opportunities has been under close public scrutiny. Both before and after the 1972 Supreme Court's decision in *Swann v. Charlotte-Mecklenburg*, the allocation of educational resources has been investigated by a number of local civic groups, including the League of Women Voters, the NAACP, as well as parent groups (Douglas, 1996; Morantz, 1996). Though Superintendent Murphy might charge the gifted education task force to "dream about what we would like to see," the designers were well aware that their work would have to stand up under public scrutiny.

A history of public scrutiny may have helped spur the implementation of the observer team that Ty Fox suggested, despite Romanoff's initial resistance. The notion that similar performances across different schools were yielding quite different designations would be hard to defend in a public forum. Establishing the observer team, in turn, has yielded a pool of trained observers and is moving the PSA in the direction of observer reliability.

School District Context

Charlotte's designers, unlike DISCOVER's, are surrounded by multiple sources of feedback about the assessment's design, implementation, and outcomes. Romanoff continues to be a school-based teacher of the gifted while revising and administering the assessment. Reid interacts with the district's 50-plus elementary school teachers of the gifted. Further, they live with the consequences of what they design in a very immediate and personal way. Udall's next door neighbor is a teacher in one of the gifted magnets. One of Romanoff's twin daughters was identified on the PSA, the other was not.

To make the assessment workable for their neighbors, colleagues, and themselves some of this feedback had to be incorporated into the assessment. As Udall said:

> I'm a coordinator in a huge school system, and if I can't make it work here, and if I can't make it practical, then it's not gonna fly.... One of the differences that I sometimes think about in terms of school system folks or university folks is exactly this issue: You know, at which point are people willing to make some pragmatic decisions...?

Several critiques of the PSA have influenced its current form. For example, Fox, Reid, and Romanoff all noted that, after the first year of the assessment, teachers asserted the new, somewhat rough-and-ready approach, selected too many youngsters who seemed unprepared for gifted education. Given this, Reid said it became necessary to "uplevel the challenge" of the assessment to identify youngsters who met the expectations of gifted teachers (see below "Curriculum-Assessment Link"). To do this, tasks that drew on notational abilities were added (e.g., sequences, functions, context, and categories). As a result, unlike DISCOVER, the PSA is no longer intelligence-fair.

Critique from within the district also facilitated the development of clear scoring procedures. Udall noted that "streamlining [the checklist] has been a large part of our responsiveness to public concern." Similarly, Romanoff stated that teachers just would not tolerate the complicated process used by DISCOVER. Reid said that those implementing the assessment pressured the designers to "make the observation process easier." As a result of dealing with this feedback, the yellow card is clear and manageable. Observers rarely said that documentation was a problem and none questioned the meaning of any of the product and process characteristics listed on the yellow card.

Though based in a school district, the PSA has also been informed by research. However, unlike DISCOVER, research on the PSA is not devoted primarily to validating the existing instrument. Rather, the research on the PSA partly helps spur revision of the assessment to make it clearer and stronger. For example, with the assistance of Professor Bob Algozzine of the University of North Carolina at Charlotte, the designers have undertaken an item analysis of the checklist. This investigation has helped the designers to eliminate checklist items that do not contribute much to the decisionmaking process. In turn, this enables the Charlotte observer team to work from a smaller and more comprehensible set of characteristics. This, alongside public critique, is taking the PSA toward clear scoring procedures.

Curriculum-Assessment Link

The force of the curriculum on the development of the PSA is akin to that noted above in the discussion of DISCOVER. In Charlotte-Mecklenburg, as in nearly all schools, the 3Rs are valued as ends and means of learning. Thus, identifying youngsters for educational programs on assessments of other strengths is likely to create a mismatch between the children selected and the education offered. As one staffer said:

> [W]e were afraid to assess children, if once identified, then we didn't have a program to meet their needs. And we were afraid ... we were biting off more than we could chew; that we better at least deal with [i.e., develop asssessments for] that which we think we could address [i.e., in the curriculum].

Similarly, Reid noted a need to have the assessment "parallel our service delivery. [If] we identify an interpersonal child, what are we doing in class for that child?"

Like Maker, Reid believed the assessment would start in familiar areas and then branch to additional intelligences. However, in Charlotte, unlike Chinle, this possibility was more limited by an extensive, pre-existing gifted program, whose students had been selected primarily on the basis of traditional tests. Teachers were accustomed to working with students selected on those assessments. Reid stated that the designers rejected linking identification to a student's reading level,

> But we have also acquiesced to the teachers' comments that in order to perform in the program, in the classes, and do the kinds of intensive research and work that is anticipated for them, they need to have some of those [language] skills. So therein lies some of the contextual clues...

Thus, while designers wanted to create more hands-on problem solving tasks -- to "get as far away from paper and pencil tasks" as feasible -- intelligence-fair tasks were at odds with the language and notational skills educators of the gifted traditionally demanded.

Alongside the gifted curriculum was the influence of the arts curriculum and arts educators. Charlotte-Mecklenburg has art and music teachers throughout the schools. In addition, in 1993 a new K-12 magnet program for the performing arts was established for anyone who wants to apply. When the new assessment was first being discussed, the art teachers voiced strong opposition to early identification in music and art. They felt such identification would undermine their goal of nurturing all students' artistic abilities.

All told, the curriculum offered by many PG teachers early on in the PSA's development, as well as the desire to maintain broad opportunities in the arts, has constrained the PSA to the three traditional areas. In addition, the development of more intelligence-fair tasks was checked by the orientation of gifted teachers toward more second-order, notationally-based classroom work.

Resources

Just as in DISCOVER, human and financial resources limited the PSA to logical-mathematical, linguistic, and spatial assessments. Because the ratio of observers to students is lower than in DISCOVER (commonly only one observer to four or five students, instead of 1:5-6), personnel costs are higher. The cost in terms of time is also high. When I observed in Charlotte, the team spent the entire school day assessing one classroom of youngsters and then several hours thereafter evaluating the students' performances. Romanoff estimated that the district spends about $170,000 a year to administer the PSA, and there is pressure to reduce this cost. Such economic forces dampen the development of assessments for additional intelligences.

Another resource limitation is that while Charlotte can identify more students than the 3.9 percent funded by the state's allotment, the district must then provide the funding for these students' participation in gifted education. As it is, because approximately 10-12 percent of the district's students are identified, financial resources are stretched thin. Assessing additional intelligences could yield additional students, which would further tax an already overburdened budget. Thus, strained resources discourage the assessment of a wider range of abilities.

Leadership

Various layers of the district leadership helped to shape the PSA. Its genesis was spurred by John Murphy, the former superintendent. According to Passe and Reid, when Murphy formed the task force on gifted education he wanted to change both the programs that were offered and the racial composition of the students served.

Udall, who became coordinator of the Program for the Gifted some months after the task force was convened, brought with her a vision for greater equity akin to that of her mentor, June Maker. She noted that "one of the last things in the world" she wanted to be involved in was a "homogeneous program ... not reflecting the diversity of the [school] system."

However, Udall, unlike Maker, was not committed to the assessment as an object in and of itself or to its imminent validation. Rather she sought to change opportunities by encouraging her staff to develop an assessment that was workable in Charlotte-Mecklenburg. Thus, when teachers in the gifted program were not happy with the complexity of the new assessment or the skills of many of the identified students, the designers addressed these concerns in ways that reduced the proportion of intelligence-fair and domain-based tasks. To make the team's work manageable also required eliminating little-used behaviors from the checklist. This streamlining is helping to foster clear scoring procedures.

In short, Charlotte's leaders maintain a vision of greater equity, but they steer a pragmatic course toward it. The resulting assessment is not drawing deeply on MI. What then accounts for the increase in identification of underserved students? Reid, Romanoff, and Udall offered similar perspectives in response to this question.

Romanoff and Reid both feel that the PSA gives observers more clues about children's potential than does a standardized test. A key source of such clues is the preassessment lessons. These employ a variety of materials and activities of the sort that children encounter on the PSA. Thus, through these lessons children are better prepared for the PSA's challenges than they are for those of a traditional, standardized test. For children from more advantaged homes, such explorations might not provide as much 'value added' (or increase the learning curve) as they might for those from less-advantaged homes. If so, as Reid stated, the preassessment lessons move "toward leveling the playing field."

The preassessment lessons may also alter teachers' views of students' capabilities, as does the DISCOVER assessment (see above). For example, Romanoff noted that teachers come away from the preassessments with comments like "'I didn't think he could do that,' or 'I didn't think he could do this.'" Thus, preassessment offers a learning experience for teachers as well as students. With new insights about students' potential from the preassessments, teachers might then provide students with more enriched classroom activities. Thus, students might actually become more competent in the areas the PSA measures by the time they take the actual assessment.

In addition, Udall noted that the preassessment lesson yields more referrals of traditionally underrepresented students to take the actual PSA. Udall believes that this higher referral rate "is *the* key to increasing identification."

Along with the preassessment, Udall said "the PSA is designed to be observational and forces people to really look at behaviors...the PSA by design creates a heightened awareness of student abilities." Romanoff and Reid felt similarly. Romanoff noted that the PSA provides observers with opportunities "to see kids solve problems using strategies, [and to] spend time with them," Romanoff said. The designers feel that because the observation extends over time and is not limited to paper-and-pencil, observers detect a range of abilities "that are indicative of intelligence, whether they are pure MI or not." Certainly, this range was obscured when Charlotte relied on standardized group tests in which students and evaluators had minimal interaction.

LOOKING INTO THE FUTURE OF THE PSA

Given the forces shaping the PSA what is the likelihood that the PSA will meet conditions it does not yet meet? With regard to the two general conditions, clear scoring procedures and observer reliability, the chances are quite good. There is work already underway, including the development of rubrics, the elimination of unnecessary checklist items, and the reliance on trained and experienced observers that supports these conditions. The chances are also good because the designers are attending to feedback from the teachers of the gifted, members of the observer team, and an outside researcher, Professor Algozzine. In addition, possible scrutiny from parents and civic groups puts pressure on the designers to make this work clear and justifiable.

Meeting the three MI-specific conditions will be a much harder task. As noted, the resources to develop and administer additional tasks and to serve youngsters identified on a broader basis are not readily available. In addition, broadening the assessment has been in conflict with the curriculum that the teachers of the gifted historically provided. It also conflicts with the goals of Charlotte's arts educators. For all these reasons, the designers of the PSA may remain restricted to the assessment of language, logical-mathematical, and spatial abilities.

In contrast, there is a somewhat better chance that the assessment will become more intelligence-fair. One countervailing force here is the increased emphasis on early attainment of reading and writing skills on the part of Charlotte's new superintendent, Dr. Eric Smith. To the extent that this emphasis filters into the PSA, the PSA may continue to require even the youngest children in the least affluent schools to represent their abilities in notational form. As noted in Chapter 3, requiring notational skills at an early age runs contrary to Charlotte's aim of enhancing equity, since poorer youngsters are likely to enter school with fewer preliteracy skills than their wealthier age-mates.

However, there are now also opportunities that lead in the direction of intelligence-fair tasks. As described above, the development of more paper-and-pencil tasks on the PSA was a response to the desire of PG teachers to identify students who could manage their curriculum's notational demands. However, Udall, Reid, and Fox noted that for the last two or more years teachers have expressed satisfaction with the abilities of the identified children. Given teachers' satisfaction, the designers of the PSA have a bit more room to experiment. They can consider introducing some new, more intelligence-fair tasks. Additionally, there is some dissatisfaction on Romanoff's part with the current set of tasks. She is working with colleagues to develop more hands-on approaches involving manipulative materials. If the assessments of logical-mathematical ability become more intelligence-fair, then, together with the existing assessments for spatial ability, some youngsters can be identified on an intelligence-fair basis.

Looking further into the future, if PG teachers continued to be satisfied with youngsters selected on more intelligence-fair measures, their curriculum might expand to address the greater range of problem-solving abilities these youngsters bring with them. Such a dynamic would make the program for the gifted into a more welcoming environment for increasingly diverse students.

As for becoming a domain-based assessment, there is both reason for optimism and pessimism. A desire for clear scoring procedures and the development of rubrics aligns with domain-based approaches, because domains are valued and judged: they have standards. On the other hand, the PSA seems committed to the Pablo® and tangram tasks as a means of detecting spatial abilities. Given this, both the new math tasks and the revised linguistic tasks would have to be domain-based for the assessment as a whole to be considered domain-based. In short, while not out of the question, reorienting the PSA around domain-based tasks is likely a long, uphill battle.

Although the PSA is not likely to meet all the MI-specific conditions, in the next year or two it can meet all the general conditions needed to make inferences about students' abilities from the assessment. The PSA has already achieved far greater equity in identification -- roughly doubling the number of African American students over the previous assessment. When the PSA meets these general conditions, then these more equitable outcomes will be justifiable not only on moral grounds but on technical grounds as well. This will support the assessment's continuity, even under scrutiny from critics of equity and critics of non-traditional assessments.

The future of the PSA has recently been given a considerable boost by changes in North Carolina state law. In January 1997, guidelines set forth by the state require each district to develop its own criteria for identification and to put these into effect by March 1988. Thus, the state's test-based method will no longer be a mandated alternative to the PSA.

Despite its promise, the future of the PSA is still not assured. To advance the PSA on the technical front, evidence needs to be gathered that it is a tool that identifies youngsters who are gifted problem solvers or "gifted" by some other definition. This will entail a range of investigations, including at least studies of observer reliability, longitudinal studies of children in the district, and comparisons of achievement scores, grades, and other outcomes by students' selected and not selected. Though some of this work is being undertaken by Algozzine and by Romanoff, leadership at or near the superintendent's level will be needed for validation to go forward. Murphy's leadership and commitment to equity in gifted education enabled the start of the PSA. Current and future superintendents' commitment will be needed for additional efforts on it.

Further, the effort to revise tasks and to ensure adequate training is quite taxing. Much of the day-to-day activity on this front has been carried out by Romanoff. Romanoff is also a doctoral student under Maker, a full-time teacher of the gifted in the district, and a mother of three children. Though she clearly is highly engaged by the work on the PSA, alongside her other responsibilities, there is the potential for burnout.

Noting the presence of these and other potential perils, I am cautiously optimistic that the PSA will be strengthened and survive. Given the state's new policy on identification, it may be that much of North Carolina will be looking to Charlotte-Mecklenburg for guidance.

MONTGOMERY COUNTY'S GIFTED MODEL PROGRAM

The effort to draw on MI to identify culturally diverse youngsters for gifted education took hold in Montgomery Knolls Elementary School, one of two schools in Montgomery County supported by Javits funding. At Montgomery Knolls, all three of the MI-specific conditions were met. It is the only one of the sites in this investigation in which it is reasonable to link the assessment to MI. As far as the general conditions are concerned, the first, "children understand the tasks," is not wholly applicable. The identification process at Montgomery Knolls was not based on tasks or specific assessment activities but rather on classroom curriculum and performances. Children's understanding of the curriculum in Montgomery Knolls was variable, as might be expected in most K-2 classrooms. The children were encouraged to do their best work in the classroom setting, and teachers were trained to provide enriched curriculum and to observe and develop students' strengths. However, as was the case in the other two sites, there was not adequate evidence that scoring procedures were clear or that the teachers/evaluators had achieved reliability in their judgments. Because all of the general conditions were not met, it is not reasonable to make inferences about students' abilities from this MI-influenced work.

To understand why the Model Program was able to meet the MI-related conditions requires looking at the forces that helped to shape it. These same forces shed light on the slender chances of the Model Program ever meeting the general conditions. They also reveal why the Model Program was not adopted elsewhere in Montgomery County. Finally, investigating these forces illuminates a question unique to this site: why didn't the MI-influenced work taking place in the school ever become a formal part of the identification process?

State Policy

Maryland, like Arizona and North Carolina, requires school districts within its borders to identify and establish services for the gifted and talented youngsters (Paynter, personal communication, February 1997). However, Maryland's state policy is the only one in which identification procedures reach well beyond the traditionally tested linguistic, logical-mathematical, and spatial realms. The state defines gifted and talented students as those with "outstanding abilities in the area of general intellectual capabilities; specific academic aptitudes; or the creative, visual or performing arts" (Maryland State Department of Education, 1983, p. 5). In line with this definition, since at least 1983, the state did not identify youngsters based on standardized tests. Thus, unlike the designers of Charlotte's PSA and DISCOVER, the designers of the Model Program were not constrained by state policy to develop their assessments around linguistic, logical-mathematical, and spatial abilities.

Instead of standardized tests, the state called for school-based committees to identify children using multiple "subjective" and "objective" indicators (Maryland State Department of Education, 1983, p. 7). These include observations of students, evaluations of their products, student auditions, interviews, biographical data, and rating scales. Since the state allows schools to consider auditions and products, identification processes can incorporate students' domain-based work. Because the state identification policy encourages the use of a range of evidence, it supports intelligence-fair approaches: students' abilities can be judged in media central to their expression, such as music, constructions, artwork, or movement. Maryland's state policy created a climate conducive to the development of domain-based and intelligence-fair identification found at Montgomery Knolls. (See Chapter 4.)

History

Alongside the favorable context created by state policy, Montgomery Knolls' own history complemented efforts to incorporate MI. In fact, one of the reasons that Starnes ultimately placed the Javits-funded effort at Montgomery Knolls was because of the potential for an MI-influenced project to work there. This potential was partly due to

Pam Prue, the school's principal (see "Leadership," below). There was also a large cohort of experienced and well-trained teachers. The PADI program had been in effect for about two years in the school. According to Donnelly Gregory, PADI's coordinator, at one point, all of the first and second grade teachers at Montgomery Knolls had been trained in PADI techniques. These are geared toward creating classroom structures and curriculum to nurture the abilities of traditionally underserved youth. In addition, Montgomery Knolls already had an enriched curriculum that would draw out children's strengths (see "Curriculum-Assessment Link," below). Given all this, the school's reach to achieve domain-based curriculum beyond the three traditionally-tested areas and to use intelligence-fair approaches was within its grasp.

School-District Context

Given that MI-related identification was achievable at least in Montgomery Knolls, why wasn't the role of MI formalized into the assessment process? This is especially odd since, as noted, Javits funding was provided partly to "confirm the validity and value of using ... multiple intelligences ... to identify gifts and provide instruction in a public school setting" (U.S. Department of Education, 1994, p. 25).

One way to understand the lack of formalization is by comparing the context of Montgomery County with that of Charlotte-Mecklenburg. In Charlotte, Superintendent Murphy and leaders in the Program for the Gifted from at least 1991 have been committed to dismantling segregation in gifted education. However, over roughly the same period of time, Montgomery's superintendent, Dr. Paul Vance, has evinced far less commitment to changing the ethnic and racial makeup of the schools and programs in his district (Eaton, 1996; Orfield, personal communication, April).[3]

Further, in Charlotte, there was a recognition among the leadership that the identification procedures were inadequate and that, by changing the identification mechanism, the composition of identified students could also be changed. In Montgomery County the picture is more mixed. There was a sense expressed by Starnes that students from minority groups were still not being detected. However, there was also a sense that the identification process was already quite good. For example, Starnes felt that even before the Model Program began, the district employed advanced concepts in identification, such as using diverse subjective and objective indicators.

In several follow-up conversations, different views emerged about why MI was never formalized in the identification. Prue said that MI enabled the Model Program teachers to "do what we have to do find out what those strengths are." Given this, formalizing its role would have been a "logical next step." However, perhaps because of the high regard for the existing identification process, the opportunity for turning MI into a formal indicator may simply have been overlooked. She noted "We were kind of locked into the system's identification process."

As Prue reflected on the question longer, she felt that Starnes may have headed in the direction of system-wide adoption had she not retired. A phone call to Starnes supported this notion. However, she also felt that the leadership in the district was indifferent to the Model Program and its aims. The leadership welcomed the additional resources, but did not invest much energy in following the actual work or spreading its methods to the 123 other elementary schools.

Another staffer said that when she raised the question of formalizing MI, about two years after the Model Program began, she was confronted with the large bureaucratic hurdles that such a change would entail:

> Who was going to deal with the whole county? Getting the county to deal with almost anything is monumental. It's huge. It's kind of bizarre: why bother doing this [i.e., developing the Model Program] if you don't want to make changes? I don't think [systemic] changes were really in some people's minds.

Because of the school district context, including lack of superintendent-level interest, bureaucratic hurdles, the regard for the county's existing identification, and possibly other reasons as well, MI simply remained an informal basis for screening committee members' advocacy. (See Chapter 4.) Since it was never formalized into a screening committee instrument, there was little incentive to develop clear scoring procedures around it, like those accompanying the district's other identification instruments. (See Chapter 4; see Appendix I.) Meeting the condition of clear scoring procedures was within reach: in many cases teachers were using observations and objects from domain-based practices, for which external criteria exist and norm group performances can be established. For example, a child's construction can be considered for form, expressiveness, composition, or other characteristics relative to children of about the same age and/or experience. (See Chapter 1.)

As for achieving reliability, the school district context, and the structure of gifted education within it, may have made this a less pressing issue for the Model Program. A comparison of Charlotte and Montgomery County sheds light on this point. In Charlotte, because gifted education and equity are prominent concerns, an assessment procedure affecting both issues could well come under scrutiny. To be able to face such scrutiny, Charlotte's designers have to address observer reliability and scoring procedures. In contrast, as noted above, within Montgomery County's school system racial imbalances are not addressed as actively and publicly. Further, since deep budget cuts in the early 1990s, the profile of gifted education has been reduced. (See Chapter 4.) In these circumstances, it is possible that school administrators, parents, or civic groups are less likely to examine the identification process itself. Thus, designers of the Model Program may not have felt a need to demonstrate reliability or clear scoring procedures around their MI-infused methods.

However, as noted in Chapter 4, relative to obervers in the other two sites, members of Montgomery Knolls' screening committee expressed more confidence in their judgments of students. This feeling was grounded in the greater amount of knowledge and information the observers brought to the table: the observers were largely teachers who had worked with the students over time. Yet, no one sought to quantify inter-observer reliability, perhaps because of the low-key role of gifted education and equity in Montgomery County.

Curriculum-Assessment Link

Within Montgomery Knolls, MI meshed comfortably with the existing curriculum and philosophy of education. Under Prue's leadership, the school was attuned to "developmentally appropriate practice." That is, teachers were already encouraged to recognize and tap children's interests and plan learning activities around these. Per Prue, MI "just fit what we were already trying to do." Alongside this philosophy, the school was already using whole language, enriched science, had school-wide music and art, and lots of hand-on learning. Much of the staff was trained in PADI, and therefore had enriched approaches to teaching, especially in the areas of science and social studies. Furthermore, the grant brought in new staff with expertise in diverse curricular areas. (See Chapter 4.) With Prue's support and leadership, teachers themselves examined and altered their practice to incorporate MI.

One of the reasons that curricular efforts that complemented MI could go forward at Montgomery Knolls, and part of the reason that the Javits-funded effort was less successful at Pine Crest, is that it did not conflict with existing assessments. The new state-wide performance assessment, and other county-wide academic measures are not administered until third grade. Because Montgomery Knolls ends in second grade, teachers there felt free to explore all the intelligences. They also felt comfortable nurturing children using intelligence-fair materials, rather than relying largely on paper-and-pencil activities.

The absence of testing pressures at Montgomery Knolls was complemented by Prue's stance toward identification. In contrast to observer team members in Charlotte, identification under Prue was not greatly influenced by children's capacity to deal with curriculum in later grades. Instead, Hylton said, "She'd be more inclined to say, 'Well, you know, this is the stuff [ability] we see. Then they [future teachers] need to address the child.'"

With its rich curriculum, the absence of state testing, and Prue's views on identification, the screening committee members could make judgments upon a wide range of students' abilities, intelligence-fair experiences, and domain-based curriculum. The teachers did not have to reshape their curriculum to meet accountability tests; they did not feel compelled to reshape their identification methods to suit the gifted curriculum of the later grades.

Resources

The effort in Montgomery County to draw on MI was endowed with many resources. First, as noted earlier, the district is considered affluent, and most of the teachers in Montgomery Knolls were highly trained. In addition, between 1990-1992, the Javits program provided $823,330 to start the Model Program at Montgomery Knolls. From 1993 through 1995, Javits awarded MCPS an additional $659,931 to further the work at Montgomery Knolls and extend the effort into Pine Crest (Barnes, personal communications, February 1997). On a per school basis, the Montgomery effort was by far the most well-funded of the three sites under investigation.

Furthermore, the Model Program did not devote resources to the development of new tasks or the creation of an observer team. Starnes noted that the county already spent a lot of money on testing. Rather than developing more testing, she wanted to find something that "we could use for ongoing assessment and instruction ... [for] finding out where the kid is and what they need."

Given that funds weren't directed toward the development of discrete assessment tasks, there were more resources that could be used for staff development and the hiring of additional personnel to enhance the learning environment. Because children were being judged on a range of performances that were infused into the curriculum on a daily basis, rather than on a separate set of relatively novel tasks, it is reasonable for them to appear more competent and for more of them to be identified.

Given that funding was not a major obstacle, at least relative to the other sites, the designers of the Model Program lost an opportunity to establish the reliability of teachers' observations and the clear scoring procedures that would justify their increased rates of identification. This, in turn, might have helped foster a formal role for MI in the county's identification procedure. Because assessment and curriculum influence each other, if MI had been formalized in gifted identification, more teachers in other parts of the county might have been spurred to generate the enriched opportunities employed in Montgomery Knolls.

Another resource-related issue may help to explain the lack of formalization of MI in Montgomery County. At least in Montgomery Knolls, MI was associated with an increase in the overall rate of identification. However, at the time of the Javits grant, the district had cut back on services for gifted education. The potential of MI to increase demands for gifted services would make it an unwelcome addition to the identification process in a time of reduced resources for gifted education.

Leadership

One of the reasons that the effort to adopt MI within Montgomery Knolls was so successful was clearly the leadership provided by its principal, Pam Prue. Literally *everyone* I talked with about the Model Program pointed to the key role that Prue played. Prue worked with teachers to help them incorporate MI into their classroom practices. She also based her teacher evaluations partly on teachers' efforts to incorporate MI. She encouraged staff to look at students in terms of their strengths, and she was noted for modelling this behavior herself with both teachers and students. As noted above, her commitment to drawing on MI stemmed in part from the complementary nature of the theory and her own philosophy of developmentally appropriate practice.

While Prue's leadership certainly enabled MI to take root within the school, there were disjunctions in the leadership above the school level. Some of these disjunctions made it difficult to draw lessons from the MI-influenced work of the Model Program. This, in turn, undermined chances of disseminating such lessons across the district -- even though the Model Program was meant to do just that.

One of the key schisms in leadership was between Dr. Starnes and Dr. Gregory. Gregory coordinated PADI for several years prior to the Model Program. Through teacher training, curriculum, and early nurturing efforts in 30 schools, PADI already enables about 25 to 30 percent of the youngsters it serves ultimately to be identified as gifted and talented. Given her experience with PADI, there is little doubt that Gregory would have been a good source of information about identification, training, and broad scale implementation issues. Unfortunately, Gregory was rarely consulted about the Javits-funded work. In fact, when I met her in late 1995, as the Javits grant was ending, she believed that MI was to be used as a tool for teachers "to heighten their awareness of [students' abilities'] and document it in some way." "I never heard it [MI] discussed as a tool for [identification]." In light of parallel efforts made by PADI "it wouldn't have made sense to me."

Since Gregory keeps a data base on all the schools she works with and has staff to analyze this data, I believe she could have pointed out the potential for ceiling effects in the identification of African American students. As noted in Chapter 4, these students were already identified at rates that were proportional to their presence in the wider school population. However, Gregory said that she was "never asked for any data to show that" there were youngsters that went undetected and unnurtured by PADI that should have been detected. As noted in Chapter 4, Gregory felt that:

> Without some structures in place, no matter how much people believe in it [MI], it's going to be difficult to point to [the MI-influenced work] as the difference that affects things that you can actually demonstrate. ... [Starnes wanted] demonstrable, significant results of this project. The unfortunate thing is that she never had any way of designing a way to look at it.

The schism between Starnes and Gregory stands in contrast to the effort by leaders in Charlotte to incorporate suggestions by a wide range of Program for the Gifted staff. The leadership in Mongtomery's Javits Program was much more like that in DISCOVER, where critique was not sought out.

Along with the disjunction within the leadership of the county's gifted and talented programs was a disjunction between the gifted and talented leadership and the superintendent's office. In Montgomery County no one I spoke with spontaneously mentioned Superintendent Vance, except to comment briefly that with his superintendency came drastic cuts in the gifted and talented budget. In later conversations, Starnes and others mentioned that Vance was happy to see the money come into the county and to be uninvolved in the project itself. In contrast, throughout Murphy's tenure in Charlotte, those in the Program for the Gifted office recognized that they had the superintendent's support. Without such interest and support, it is hard to imagine how system-wide changes in identification could occur.

LOOKING INTO THE FUTURE OF THE MODEL PROGRAM

Of the three efforts to draw on MI, the future of the Model Program's work is clearest: There is probably no systemic future for the program because it has no institutional champion in the county: Starnes has retired; Prue was promoted to director of the county's Division of Early Childhood Services; and the new principal of Montgomery Knolls has not been involved in sustaining the work on MI accomplished under Prue's leadership. The only slim possibility for the Model Program's influence to continue rests with Pam Sobel, the principal of Pine Crest Elementary. As mentioned in Chapter 4, by the end of the second Javits grant Sobel believed the teachers at her school were at last ready to embark more fully on MI-infused education (MCPS, 1996, Appendix N). Unfortunately, these teachers are also operating under a new state-wide testing framework, which they perceive as somewhat at odds with MI. Furthermore, they no longer have grant support to provide training, curriculum development, or other resources needed to apply the theory. Given these circumstances, the chance of MI taking hold in Pine Crest and spreading outward from there are minimal.

Hylton points out that, despite the lack of systemic change in the identification procedures, many teachers have been influenced by the Javits work. As a teacher trainer and curriculum developer, Hylton has incorporated MI into her training. Prue reports MI now infuses the efforts of the Division of Early Childhood Services, which she directs. Here as in all the sites I have visited for this and other research on MI, the theory clearly had a powerful impact on the adults (Kornhaber, 1994; Kornhaber & Krechevsky, 1995). As Prue said:

> [MI fostered a] transformation of how we talked about children, and the elimination of this deficit model. I'm not sure that they [students] weren't always showing it [their strengths], and we just missed it because of our blinders. But I think it transformed our teaching and took the blinders off.

However, unlike the other sites, the Model Program's effort has left neither systemic change in the identification procedures nor more equitable access to enriched curriculum (whether or not such access could be rationally tethered to the identification method). Given the curricular links at Montgomery Knolls, the high level of resources that were available, and the synchrony between MI-influenced approaches and state-wide identification policies, there was a solid opportunity to make systemic change in the identification. This opportunity has passed.

CONCLUSION

As this chapter highlights, the three sites vary with regard to the conditions they meet not only because their designs differ (see Chapters 2-4), but also because of differences in the contexts in which those designs were developed and implemented. State policy, local history, institutional setting, resources, leadership, and the link between assessment and curriculum shaped the work and will continue to influence its future.

While the assumptions underlying these three assessment efforts are revolutionary, the realities of implementing them in context makes the actual work evolutionary. With regard to the MI-specific conditions, DISCOVER was constrained from the start by state policy governing the realms in which identification should be made. The work was also underfunded, making it difficult to develop and administer a broader set of tasks; state funding for gifted education also undermined a broad assessment. In Charlotte, state identification practices and funding also limited the development of tasks to tap the range of students' strengths. In addition, teachers of the gifted assumed a student body strong in notational skills. These teachers, and their curricula, could not be instantly modified. Therefore, as one designer noted, the new assessment was not undertaken with "a slash and burn tactic." At Montgomery Knolls, there was a leader and a set of school practices that made MI a comfortable fit. However, formalizing MI-influenced practices and exporting these to the rest of Montgomery County was hindered by beliefs that the existing assessment program was strong and by disjunctions within the leadership.

As with the MI-specific conditions, the ability of sites to meet the general conditions was also shaped by the context of the work. For example, DISCOVER's leadership maintains that scoring procedures are already clear and reliability has been achieved. The university environment also diminishes critique from teachers, administrators, and parents, allowing the designers to maintain unwieldy scoring procedures and inadequate observer training. In Charlotte-Mecklenburg, issues of equity and education are publicly debated. Given this, the designers there were inclined to use trained observers and are working on developing clear scoring procedures and reliability. In the Model Program, disjunctions in leadership undermined formalization of the MI-infused approach. This left little concrete basis upon which to establish clear scoring procedures or observer reliability.

While the assessments' evolution yielded adaptations that met some, but not all the eight conditions, unlike biological evolution (at least the scientific view of it), there are thinking beings behind these designs and the designs' revisions. Though I have attempted to look into these programs' futures, this investigation or other features of the surrounding context could alter the designers' thinking and their assessments. In part, this work is aimed at doing so. Because there are still leaders and funding streams supporting DISCOVER and Charlotte's PSA, there are opportunities to enhance the important work that has been done there. In both places, clear scoring procedures and observer reliability need to be developed and/or demonstrated. DISCOVER observers need to have more uniform training and deeper experience to carry out their work. These general conditions need to be met if designers' claims of enhanced equity are to be associated with their assessment procedures.

Beyond providing feedback to the sites, this work has also been aimed at providing a framework, in the form of the eight conditions, for other educators who are contructing, or contemplating, identification processes that draw on MI (or other theories). My hope is that this framework helps others to devise equitable, powerful, and persuasive identification efforts.

Reflecting on the work in these three sites can also be useful to policymakers who are considering using MI in identification for gifted programs or in other types of assessments. The work investigated here points to some characteristics policymakers might look for and encourage in the development of such assessments:

1. Efforts to use MI should clearly articulate what the theory will look like in practice.

The theory of multiple intelligences, unlike most ideas used in education reform, was not originally associated with any school-based practices (Kornhaber, 1994; Kornhaber & Krechevsky, 1995). To illustrate, the index to Frames of Mind (Gardner, 1983), in which MI was first introduced, does not have entries for "curriculum," "pedagogy," "teachers," or "students," let alone, "assessment," or "tests." It was initially put forward as a theory about the human mind, not a theory about education (Gardner, 1995).

Since its first appearance, many people have generated ideas about how to devise curriculum, assessment, and pedagogy incorporating MI. I have used three MI-specific conditions as fundamental to assessments that draw on the theory: First, that the assessment extends beyond the three abilities that are traditionally measured (i.e., linguistic, logical-mathematical, and spatial). Second, the assessment should be intelligence-fair, so that youngsters' abilities can be tapped without undue reliance on paper-and-pencil activities or verbal skill. Finally, the assessments should be domain-based, or built on practices that matter in the surrounding culture and for which standards exist.

Given that there is no definitive or exhaustive list of such conditions, it is certainly possible to argue for alternatives to the three I have suggested. However, policymakers should see some clearly articulated set of MI-related practices before funding any "MI" programs or assessments, because the theory itself does not provide this. With regard to MI, policymakers should look for something less "sexy" -- an adjective applied to MI by one site's grantwriters -- and more concrete.

2. Efforts to draw on MI in assessment must consciously select from key general principles of assessment.

Assessments incorporating MI cannot completely abandon the preceding history and practice of assessment. Nor is it entirely desirable to do so. Drawing on some standard works in the field of assessment (e.g., Cronbach, 1990; Sattler, 1992), I have suggested five general conditions that are useful to incorporate in assessments drawing on MI: students understand the tasks; students are encouraged to do their best work; the observers are adequately trained; there are clear scoring procedures; the observers' judgments are reliable. I believe that meeting these conditions will make the identification process publicly defensible.

It is possible to argue for other conditions, but whatever the conditions are, these need to be clearly stated and justified. Such clear statements may help prevent a mix of useful and potentially unnecessary or contradictory practices from being implemented. For example, given that MI argues that an ability is manifested in a cultural practice or domain, efforts to maintain novelty might be reconsidered. Policymakers who are considering efforts to support MI-infused assessments should look for designs that have detailed the traditional test characteristics that are being incorporated and how and why these characteristics are being used.

3. Beyond the actual design of the identification process, the assessment needs to be situated in an enabling home base.

This last chapter highlights how the context of each of the three identification efforts influenced their form and viability. Policymakers looking to support the development and systemic dissemination of such assessments should think in terms of contexts with enabling characteristics. At least two enabling characteristics come through in this investigation.

First, the site should have built-in, unavoidable, and public sources of critical feedback. In order to devise assessments that are as technically sound as possible, designers need to confront what is *not* working: the directions that students don't understand, the observers who are confused about the scoring, the overwhelming reliance on paper-and-pencil, the unwieldiness of the assessment instrument. If the designers must hear such critique, they will be more likely to address it (or turn over responsibilty for the design to others who can address it). Having many sources of unavoidable and public critical feedback helped push the PSA into its current form. The absence of these features renders DISCOVER less defensible. Similarly, because gifted education and equity were far less compelling topics in Montgomery County, the work at Montgomery Knolls went on with too little critique and public evaluation. In the end, little of the good work that happened in that school actually survived. Critique, rather than cushiness, will spur a resilient and technically adequate assessment.

A second enabling characteristic is the involvement of a group of schools, or perhaps a district, rather than one or two schools. Especially in a large district, work that takes place in a single school is less likely to attract the attention or commitment of district leaders. In addition, the more students and parents affected, the more likely the work will get attention (including the critique needed to shore it up). Conversely, work in a single school is simply too vulnerable to extinction. For example, the strong work that pervaded Montgomery Knolls began to whither as soon as the principal left. In addition, it is harder to demonstrate the nature of the benefit in a single school: to what extent was this good work related to the principal, to the changed program, to a Hawthorne effect, or something else? For policymakers looking to develop a more equitable method of identifying underrepresented youngsters for gifted education, this investigation suggests that attempts in many schools and attempts carried out in a public spotlight are likely to fare better than smaller efforts lacking public scrutiny.

It is worth noting for policymakers that Charlotte's efforts suggest that devising and implementing assessments aimed at enhancing equity appear to be economically feasible. Further, the cost of implementing such assessments elsewhere should decrease, given the concrete groundwork undertaken by the three sites and given this framework of eight conditions for developing, analyzing, and modifying assessments.

It was with the aims of enhancing existing work, creating a useful framework, and offering some suggestions to policymakers that I began this dissertation. However, in the course of conducting this research, I also came to some greater understanding of my own motivations. Sharing drafts with the sites and receiving some of their critique made me question my aims. Wasn't it enough that each of these sites had, at least for some period of time, improved youngsters' access to more challenging curriculum? In light of all the social inequities undergirded partly by standardized testing, wasn't such enhanced access justification enough?

I responded first on logical grounds: the answer would be an unqualified yes if my question had been: Do these assessments enhance equity in identification for gifted education? However, this was not my question. Instead, as described in Chapter 1, I sought to understand how MI theory is being used to identify poor and minority elementary students for gifted education. This inquiry branched into two questions: is it reasonable to associate increases in the proportion of identified students from poor and minority populations with these assessments? (The analysis of general conditions shed light on this). Second, is it reasonable to associate these assessments with MI theory? (The analysis of the MI-specific sought to illuminate this issue).

Beyond the logic of all of this, I still had some doubts on the moral plane: Shouldn't I leave undisturbed the foundations upon which these assessments rest? Even if the sites did not meet all the conditions, since they were identifying more traditionally underserved youngsters, wasn't the work nevertheless alright?

I have come to the conclusion that the work is defensible as an interim step: ultimately, the general conditions needed to make inferences about students should be met. There are at least two reasons for this. On the political front, the outcomes from such identification processes are just too vulnerable without observer reliability, observer training, and clear scoring procedures: These are essential elements in any assessment that is used to allocate educational opportunities. Those who are not in favor of equity or new approaches to identification could easily swipe away support for assessments that did not meet these fundamental conditions. The goal of the three sites' work -- enhancing educational equity for underserved youth -- is too important to allow the methods or technical foundations to collapse under modest scrutiny.

Finally, moral and philosophical perspectives also lead me to see the importance of reaching for greater technical adequacy for these or other assessments aimed at equity. If one truly holds that talent and ability exist abundantly in children across economic and racial continua, then there should be little hesitation about asking new or untraditional assessments to meet at least the five general conditions used here. Assessments can be devised which meet these five criteria *and* enhance equity. Charlotte-Mecklenburg's PSA, which is well on the way to meeting these conditions, supports the feasibility of doing so.

However, meeting these conditions and achieving greater equity likely requires a reorientation away from some other practices associated with traditional testing. Some pieces of this reorientation were present in all the sites: Assessments will likely need to draw upon more data than the single test administrations schools and districts are accustomed to using. Identification may need to meld assessment and learning opportunities in order to draw out evidence of strengths in children across different groups. The identification process may need to serve double duty. That is, it may need to be both a tool to identify youngsters' abilities and a tool to educate adults about how to nurture these abilities. Identification practices may also need to draw upon at least intelligence-fair practices associated with MI: using such measures will enable young students to demonstrate their abilities, whether or not they have been raised in literacy-rich environments.

Inequitable identification may still occur in assessments that incorporate untraditional procedures *and* that meet the five general conditions. Alongside other information drawn from well-grounded methods and arguments, such results feed revolutionary questions: is the educational setting as rich for poor and minority youngsters as it is for white and affluent ones? What else do we need to do as educators and citizens to ensure that young students' strengths are recognized and nurtured?

1. Arizona allows districts to use the assessment as a pilot measure to identify youngsters, as long as one of the nationally normed and approved tests is also administered (Stahl, personal communication, March, 1997).

2. A newer round of assessments which does expand beyond the traditional three areas is being developed for younger children. DISCOVER I assessments, devised in the mid- and late-1980s, also drew on more of the intelligences. However, as Nielson (personal communication, February 18, 1997) noted, these were individually administered, labor intensive, and not appropriate for most elementary age students.

3. According to Starnes, the retired director of the county's gifted and magnet programs, Vance had been quite committed to equity in gifted education when he served as one of four "area superintendents" for Montgomery County. Then, "Dr. Vance really worked hard on that, but as superintendent, he seems less effective" in advancing equity and desegregation. Starnes believed that over the years Vance was influenced by the county's boards of education, many of which did not see desegregating educational opportunities as a geniuine priority.

REFERENCES

Adams, C.M., & Callahan, C.M. (1994, Fall). Examining a tool for assessing multiple intelligences. The National Research Center on the Gifted and Talented Newsletter, pp. 6-7. Storrs, CT: The National Research Center on the Gifted and Talented, The University of Connecticut.

Arizona State Department of Education. (1992). Mandatory services for gifted students in Arizona. Phoenix: Arizona State Department of Education.

Baldwin, A.Y. (1994). The seven plus story: Developing hidden talent among students in socioeconomically disadvantaged environments. Gifted Child Quarterly, 38(2), 80-84.

Baska, L. (1986). Alternatives to traditional testing. Roeper Review, 8(3), 181-184.

Bloom, B. (Ed.). (1985). Developing talent in young people. New York: Ballantine Books.

Borland, J.H. (1986). IQ tests: Throwing out the bathwater, saving the baby. Roeper Review, 8(3), 163-167.

Borland, J.H. (1989). Planning and implementing programs for the gifted. New York: Teachers College Press.

Bradsher, K. (1994, November 27). Dragging banks to middle of nowhere. The Houston Chronicle. Business Section, p. 5.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32-42.

Callahan, C.M., & McIntire, J.A. (1994). Identifying outstanding talent in American Indian and Alaska Native students. Washington, DC: U.S. Department of Education.

Cassidy, J., & Hossler, A. (1992). State and federal definitions of the gifted: An update. Gifted Child Quarterly, 15(1), 46-53.

Ceci, S.J. (1990). On intelligence... more or less: A bio-ecological treatise on intellectual development. Englewood Cliffs, NJ: Prentice Hall.

Charlotte-Mecklenburg Schools. (1994a). Preliminary findings and implications of Project START. Charlotte, NC: Program for the Gifted, Charlotte-Mecklenburg Schools.

Charlotte-Mecklenburg Schools. (1994b). Vital links. Charlotte, NC: Program for the Gifted, Charlotte-Mecklenburg Schools.

Charlotte-Observer (1995, October 26). 1994-1995 Charlotte-Mecklenburg Schools Annual Report Card: Building a Legacy of Excellence. Special supplement.

Coleman, M.R., & Gallagher, J.J. (1995). State identification policies: Gifted students from special populations. Roeper Review, 17(4), 268-275.

Copenhaver, R.W., & McIntyre, D.J. (1992). Teachers' perception of gifted students. Roeper Review, 14(3), 151-153.

Cronbach, L.J. (1989). Construct validation after thirty years. In R.L. Linn (Ed.), Intelligence: Measurement, theory, and public policy. Urbana and Chicago: University of Illinois Press.

Cronbach, L.J. (1990). Essentials of psychological testing. New York: Harper & Row.

Daiute, C. (Ed.). (1993). The development of literacy through social interaction. San Francisco: Jossey-Bass.

Davis, G. & Rimm, S. (1989). Education of the gifted and talented (3d edition). Englewood Cliffs, NJ: Prentice-Hall.

Douglas, D. (1995). Reading, writing, and race. Chapel Hill, NC: University of North Carolina Press.

Eaton, S. (1996). Slipping toward segregation: Local control and eroding desegregation in Montgomery County. In G. Orfield, S. Eaton, and the Harvard Project on School Desegregation, Dismantling desegregation: The quiet reversal of Brown v. Board of Education. New York: New Press.

Feldman, D.H. (1986). Nature's gambit: Child prodigies and the development of human potential. New York: BasicBooks.

Fine, M. (1991). Framing dropouts: Notes on the politics of an urban public high school. Albany, NY: State University of New York Press.

Folio Corporation (1996). FolioVIEWS Infobase Manager 3.1. Provo, Utah: Folio Corporation.

Ford, D.Y. (1994). The recruitment and retention of African-American students in gifted education programs: Implications and recommendations. Storrs, CT: The National Research Center on the Gifted and Talented.

Ford, D.Y. (1995). Desegregating gifted education: A need unmet. Journal of Negro Education, 64(1), 52-62.

Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the "burden of 'acting white.'" The Urban Review, 18(3), 176-203.

Frasier, M. (1980). Programming for the culturally diverse. In J.B. Jordan & J. A. Grossi (Eds.), An administrator's handbook on designing programs for the gifted and talented. Reston, VA: Council for Exceptional Children.

Frasier, M. (1989a). Poor and minority students can be gifted, too! Educational Leadership, 46(6), 16-18.

Frasier, M. (1989b). Identification of gifted black students: Developing new perspectives. In C.J. Maker, & S.W. Schiever (Eds.), Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

Frasier, M., Garcia, J.H., & Passow, A.H. (1995). A review of assessment issues in gifted education and their implications for identifying gifted minority students. Storrs, CT: The National Research Center on the Gifted and Talented.

Frasier, M.M., & Passow, A.H. (1994). Toward a new paradigm for identifying talent potential. Storrs, CT: The National Research Center on the Gifted and Talented.

Frederickson, J., & Collins, A. (1989). A systems theory of educational testing. Educational Researcher, 18(9), 27-32.

Gagné, F. (1994). Are teachers really poor talent detectors? Comments on Pegnato and Birch's (1959) study of the effectiveness and efficiency of various identification techniques. Gifted Child Quarterly, 38(3), 124-126.

Gagné F., Bélanger, J., & Motard, D. (1993). Popular estimates of the prevalence of giftedness and talent. Roeper Review, 16(2), 96-98.

Gallagher, J.J. (1994). Current and historical thinking on education for gifted and talented students. In P. Ross (Ed.), National excellence: A case for developing America's talent. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Gallagher, J.J., & Courtright, R.D. (1986). The educational definition of giftedness and its policy implications. In R.J. Sternberg, & J.E. Davidson (Eds.), Conceptions of giftedness. New York, NY: Cambridge University Press.

Gallagher, J.J. & Weiss, P. (1979). The education of gifted and talented students. Occasional Paper 27. Washington, D.C.: Council for Basic Education.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. New York: BasicBooks.

Gardner, H. (1985). Reintroducing Frames of mind. (Introduction to the paperback edition of Frames of mind). New York: BasicBooks.

Gardner, H. (1991a). Assessment in context: The alternative to standardized testing. In B.R. Gifford & M.C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement, and instruction. Boston: Kluwer.

Gardner, H. (1991b). The unschooled mind. New York: BasicBooks.

Gardner, H. (1993a). The relation of intelligence to other valued human capacities. In H. Gardner, Multiple intelligences: The theory in practice. New York: BasicBooks.

Gardner, H. (1993b). Questions and answers about multiple intelligences theory. In H. Gardner, Multiple intelligences: The theory in practice. New York: BasicBooks.

Gardner, H. (1994). Intelligences in theory and practice: A response to Elliot W. Eisner, Robert J. Sternberg, and Henry M. Levin. Teachers College Record, 95(4), 576-583.

Gardner, H. (1995). Reflections on multiple intelligences: Myths and messages. Phi Delta Kappan, 77(3), 200-209.

Gardner, H. (in press). Are there additional intelligences? The case for the naturalist, spiritual, and existential intelligences. In J. Kane (Ed.), Education, information, and transformation. Englewood Cliffs, NJ: Prentice-Hall.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school. Educational Researcher, 18, 4-10.

Gardner, H., Kornhaber, M., & Wake, W. (1996). Intelligence: Multiple perspectives. Ft. Worth, TX: Harcourt Brace College Publishers.

Garrison, L. (1989). Programming for the gifted American Indian student. In C.J. Maker & S. Schiever (Eds.), Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

Gifford, B.R. (1989). The allocation of opportunities and the politics of testing: A policy analytic perspective. In B.R. Gifford (Ed.), Test policy and the politics of opportunity allocation: The workplace and the law. Boston: Kluwer.

Gregory, D., Starnes, W., & Blaylock, A. (1990). Finding and nurturing potential giftedness among black and Hispanic students. Rockville, MD: Montgomery County Public Schools.

Griffiths, S. (n.d.) The inter-observer reliability of the DISCOVER Problem Solving

Assessment. Tucson: Department of Special Education and Rehabilitiation, University of Arizona. Unpublished paper.

Gruber, H. (1986). The self-construction of the extraordinary. In R.J. Sternberg & J.E. Davidson (Eds.), Conceptions of giftedness. New York: Cambridge University Press.

Haney, W. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. Phi Delta Kappan, 70(9), 683-687.

Harris, J.H., & Ford, D.Y. (1991). Identifying and nurturing the promise of gifted black American children. Journal of Negro Education, 60(1), 3-18.

Hartigan, J.A., & Wigdor, A.K. (Eds.). (1989). Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery. Washington, D.C. National Academy Press.

Hartley, E.A. (1991). Through Navajo eyes: Examining differences in giftedness. Journal of American Indian Education. 31(1), 53-63.

Heath, S.B. (1983). Ways with words. New York: Cambridge University Press.

Herrnstein, R.J., & Murray, C. (1994). The bell curve. New York: Free Press.

High, M, & Udall, A. (1983). Teacher rating of students in relation to ethnicity of students and school ethnic balance. Journal for the Education of the Gifted, 6(3), 154-66.

Howard, J., & Hammond, R. (1985, September 9). Rumors of inferiority: The hidden obstacles to black success. The New Republic, 3686, 17-21.

Jackson, N.E., & Butterfield, E.C. (1986). A conception of giftedness designed to promote research. In R.J. Sternberg & J.E. Davidson (Eds.), Conceptions of giftedness. New York: Cambridge University Press.

Jensen, A. (1980). Bias in mental testing. New York: The Free Press.

Johnson, S., Starnes, W., Gregory, D., & Blaylock, A. (1985). Program of assessment, diagnosis, and instruction (PADI): Identifying and nurturing potentially gifted and talented minority students. Journal of Negro Education, 54(3), 416-430.

Kaplan, R.M., & Sacuzzo, D.P. (1993). Psychological testing: Principals, applications, and issues: Pacific Grove, CA: Brooks/Cole.

Kaufman, A.S., & Harrison, P.L. (1986). Intelligence tests and gifted assessment: What are the positive? Roeper Review, 8(3), 154-159.

Kitano, M.K., & Kirby, D.F. (1985). Gifted education: A comprehensive view. Boston: Little, Brown and Company.

Knox, R.A. (1995, November 5). Brainchild. Boston Globe Magazine, pp. 22-23, 38-42, 45-48.

Kolb, K.J. & Jussim, L. (1994). Teacher expectations and underachieving gifted children. Roeper Review, 17(1), 26-30.

Kornhaber, M. (1994). The theory of multiple intelligences: Why and how schools use it. Cambridge, MA: Harvard Graduate School of Education. Unpublished qualifying paper.

Kornhaber, M. & Gardner, H. (1993). Varieties of excellence: Identifying and assessing children's talents. New York: The National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.

Kornhaber, M., & Krechevsky, M. (1995). Expanding definitions of learning and teaching: Notes from the MI underground. In P.W. Cookson & B. Schneider (Eds.), Transforming schools. New York: Garland Publishing.

Kozol, J. (1991). Savage inequalities. New York: Crown Publishers.

Krechevsky, M. (1991). Project Spectrum: An innovative assessment alternative. Educational Leadership, 48(5), 43-49.

Krechevsky, M. (1992). Montgomery Knolls. Cambridge, MA: Harvard Graduate School of Education. Unpublished fieldnotes.

Krechevsky, M. (1994). Project Spectrum: Preschool assessment handbook. Cambridge, MA: President and Fellows of Harvard College.

Kulik, J.A. (1992). An analysis of the research on ability grouping: Historical and contemporary perspectives. Storrs, CT: The National Research Center on the Gifted and Talented.

Kulik, J.A. & Kulik, C-L.C. (1991). Ability grouping and gifted students. In N. Colangelo & G.A. Davis (Eds.), Handbook of gifted education (pp. 178-196). Boston: Allyn & Bacon.

Leibowitz, D. & Starnes, W. (1993). Unmasking young children's gifts. Gifted Child Quarterly, 16(5), 28-32.

Levin, H. (1994). Multiple intelligence theory and everyday practices. Teachers College Record, 79(4), 49-55.

Madaus, G. & Kelleghan, T. (1993). The British experience with 'authentic' testing. Phi Delta Kappan, 74(6), 458-59, 462-63, 466-69.

Maker, C.J. (1992). Intelligence and creativity in multiple intelligences: Identification and development. Educating Able Learners, 17(4), 12-19.

Maker, C.J. (1993). Creativity, intelligence, and problem solving: A definition and design for cross-cultural research and measurement related to giftedness. Gifted Education International, 9, 68-77.

Maker, C.J. (1994). Authentic assessment of problem solving and giftedness in secondary school students. Journal of Secondary Gifted Education, 6(1), 19-29.

Maker, C.J., Nielson, A., & Rogers, J. (1994). Giftedness, diversity, and problem-solving. Teaching Exceptional Children, 27(1), 4-19.

Maker, C.J., Rogers, J., & Nielson, A. (1995). The DISCOVER Process (Grades 3-5). Tucson, AZ: Unpublished set of instructions.

Maker, C.J., & Schiever, S.W. (Eds.). (1989). Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

Marland, S. (1971/1972). Education of the gifted and talented. Report to the Congress of the United States by the U.S. Commissioner of Education. Washington, DC: Government Printing Office.

Maryland State Department of Education (1983). Criteria for Excellence: Gifted and talented program guidelines. Baltimore, MD: Maryland State Department of Education.

McDaniel, T.R. (1993). Education of the gifted and the excellence-equity debate: Lessons from history. In C.J. Maker & D. Orzechowski-Harland (Eds.), Critical issues in gifted education: Programs for the gifted in the regular classroom. Austin, TX: Pro-Ed.

Mercer, J.R., & Lewis, J.F. (1978). Using the system of multicultural pluralistic assessment (SOMPA) to identify the gifted minority child. In A.Y. Baldwin, G.H. Gear, & L.J. Lucito (Eds.), Educational planning for the gifted: Overcoming cultural, geographic and socioeconomic barriers. Reston, VA: The Council for Exceptional Children.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational measurement (3d ed., pp. 13-103). New York: American Council on Education and Macmillan.

Metis Associates. (1994). Community School District 18 Jacob K. Javits Gifted and Talented Javits 7+ Program. 1993-1994 Final Evaluation Report. New York: Metis Associates.

Mickelson, R.A. (1990). The attitude-achievement paradox among black adolescents. Sociology of Education, 63(1), 44-61.

Mitchell, P.B. (1994). State policy issues in the education of gifted and talented students. In P. O'C. Ross (Ed.), National excellence: A case for developing America's talent. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Montgomery County Public Schools (1987a). Procedures for selection of elementary students to participate in gifted and talented programs. Rockville, MD: Montgomery County Public Schools Office of Instruction and Program Development/Gifted and Talented Unit.

Montgomery County Public Schools (1987b). MCPS handbook for implementation of procedures for selection of elementary students to participate in gifted and talented programs. Rockville, MD: Montgomery County Public Schools Office of Instruction and Program Development/Gifted and Talented Unit.

Montgomery County Public Schools. (n.d.). Parent PADI Fact Sheet. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1989). Early identification and instruction of underserved young gifted students: A model program. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1992). Multiple intelligences: A framework for student and teacher change. A proposal submitted to the U.S. Department of Education Office of Educational Research and Improvement. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1993). Multiple intelligences: A framework for student and teacher change (continuation application). A proposal submitted to the U.S. Department of Education Office of Educational Research and Improvement. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1996a). Schools at a glance. Rockville, MD: Montgomery County Public Schools Department of Educational Accountability.

Montgomery County Public Schools. (1996b). Policy: Gifted and talented education. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1996c). Regulation: Gifted and talented education. Rockville, MD: Montgomery County Public Schools.

Montgomery County Public Schools. (1996d). Multiple intelligences: A framework for student and teacher change. (Final report to the Javits Program). Rockville, MD: Montgomery County Public Schools.

Morantz, A. (1996). Desegregation at risk: Threat and reaffirmation in Charlotte. In G. Orfield, S. Eaton, and the Harvard Project on School Desegregation, Dismantling desegregation: The quiet reversal of Brown v. Board of Education. New York: New Press.

Morris, P. (1977). The bell shaped pitfall. In P.L. Houts (Ed.), The myth of measurability. New York: Hart Publishing Company.

Mullis, I.V.S., Campbell, J.R., & Farstrup, A.E. (1993). NAEP 1992 reading report card for the nation and the states. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Mullis, I.V.S, Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). NAEP 1992 mathematics report card for the nation and the states. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Natriello, G.M., McDill, E.L., & Pallas, A.M. (1990). Schooling disadvantaged children. New York: Teachers College Press.

North Carolina Department of Public Instruction (1996). Procedures governing programs and services for children with special needs. Raleigh, NC: North Carolina State Board of Education.

Oakes, J. (1985). Keeping track. New Haven, CT: Yale University Press.

O'Connell, M. (1992, November 22). TUSD program lacks diversity, officials say. Arizona Star, Section B, pp. 1-2.

Ogbu, J. (1978). Minority education and caste. New York: Academic Press.

Passow, A.H., & Rudnitski, R.A. (1993). State policies regarding education of the gifted as reflected in legislation and regulation. Storrs, CT: The National Research Center on the Gifted and Talented.

Pegnato, C.W., & Birch, J.W. (1959). Locating gifted children in junior high schools: A comparison of methods. Exceptional Children, 25, 300-304.

Perkins, D. (1995). Outsmarting IQ: The science of learnable intelligence. New York: Free Press.

Perrine, J. (1989). Situational identification of gifted Hispanic students. In C.J. Maker & S.W. Schiever (Eds.), Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

Pfeiffer, A.B. (1989). Purpose of programs for gifted and talented and highly motivated American Indian students. In C.J. Maker & S.W. Schiever (Eds.), Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

Plucker, J., Callahan, C., & Tomchin, E.M. (1996). Wherefore art thou, multiple intelligences? Alternative assessments for identifying talent in ethnically diverse and low income students. Gifted Child Quarterly, 40(2), 81-92.

Reid, C., Udall, A., Romanoff, B., & Algozzine, B. (in press). Comparison of traditional and alternative assessment criteria. Article for Gifted Child Quarterly.

Renzulli, J.S. (1978). What makes giftedness? Reexamining a definition. Phi Delta Kappan, 60, 180-184.

Renzulli, J.S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R.J. Sternberg & J.E. Davidson (Eds.), Conceptions of giftedness. New York: Cambridge University Press.

Renzulli, J.S., Hartman, R.K., & Callahan, C.M. (1980). Teacher identification of superior students. In J.S. Renzulli & E. P. Stoddard (Eds.), Gifted and talented education in perspective. Reston, VA: Council for Exceptional Children.

Renzulli, J.S., & Reis, S.M. (1991). The reform movement and the quiet crisis in gifted education. Gifted Child Quarterly, 35(1), 26-35.

Resnick, L. (1987). Learning in school and out. Educational Researcher, 16(9), 13-20.

Resnick, L. (1991). Shared cognition: Thinking as social practice. In L. Resnick, J.M. Levine, & S.D. Teasley (Eds.), Perspectives on socially shared cognition. Washington, DC: American Psychological Association.

Resnick, L., Levine, J.M., & Teasley, S.D. (1991). (Eds.), Perspectives on socially shared cognition. Washington, DC: American Psychological Association.

Robinson, N.M., & Chamrad, D.L. (1986). Appropriate uses of intelligence tests with gifted children. Roeper Review, 8(3), 160-163.

Roedell, W.C., Jackson, N.W., & Robinson, H.B. (1980). Gifted young children. New York: Teachers College Press.

Rogers, K.B. (1991) The relationship of grouping practices to the education of the gifted and talented learner. Storrs, CT: The National Research Center on the Gifted and Talented.

Rogoff, B. (1990) Apprenticeship in thinking: Cognitive development in social context. New York and Oxford: Oxford University Press.

Rogoff, B., & Lave, J. (1984). Everyday cognition: Its development in social context. Cambridge, MA: Harvard University Press.

Romanoff, B. (n.d.). A comparison of consistency decisions using traditional and nontraditional methods for identifying gifted and talented students. Tucson: University of Arizona. Unpublished paper written in 1996.

Ross, P. O'C. (1994). (Ed.), National excellence: A case for developing America's talent. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education.

Sattler, J.M. (1992). Assessment of children (Rev. 3d ed.). San Diego: Sattler.

Schiever, S. (1991). A comprehensive approach to teaching thinking. Boston: Allyn & Bacon.

Schmidt, P. (1993, May 26). Seeking to identify the gifted among L.E.P. students. Education Week, 12(35), 1, 12-13.

Schön, D. (1983). The reflective practitioner: How professionals think in action. New York: BasicBooks.

Shepard, L. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), Review of Research in Education, 19, 405-450.

Shore, B.M., Cornell, D.G., Robinson, A., & Ward, V.S. (1991). Recommended practices in gifted education: A critical analysis. New York: Teachers College Press.

Slavin, R.E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. Contemporary Educational Psychology, 21(1), 43-69.

Simon, H. (1979). Models of thought. New Haven, CT: Yale University Press.

Smutny, J. (1996). The underserved young gifted child: Status of programs and research. Gifted Education Press Quarterly, 10(1), 2-6.

Snow, C. (1991). Unfulfilled expectations: Home and school influences on literacy. Cambridge, MA: Harvard University Press.

Starnes, W. (1993). A model program for identifying underserved gifted students. In C.M. Callahan, C.A. Tomlinson, & Pizzat, P.M. (1993). Contexts for promise: Charlottesville, VA: University of Virginia.

Steele, C., & Aronson (1995). Stereotype threat and the intellectual performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797-811.

Sternberg, R.J. (1985). Beyond IQ. New York: Cambridge University Press.

Sternberg, R.J. (1986a). A triarchic theory of intellectual giftedness. In R.J. Sternberg & J.E. Davidson (Eds.), Conceptions of giftedness. New York: Cambridge University Press.

Sternberg, R.J. (1986b). Identifying the gifted through IQ: Why a little bit of knowledge is a dangerous thing. Roeper Review, 8(3), 143-147.

Sternberg, R.J. (1988). The triarchic mind. New York: Viking.

Sternberg, R.J., & Davidson, J.E. (1986). Conceptions of giftedness: A map of the terrain. In R.J. Sternberg & J.E. Davidson (Eds.), Conceptions of giftedness. New York, NY: Cambridge University Press.

Sternberg, R.J., & Wagner, R.K. (1993). The g-ocentric view of intelligence and job performance is wrong. Current Directions in Psychological Science, 2(1), 1-5.

Stiggins, R.J. (1994). Student-centered classroom assessment. NY: Merrill.

Swisher, K., & Tonemah, S.A. (1991). Preface. Journal of American Indian Education, 31(1), 1-2.

Tannenbaum, A.J. (1983). Gifted children: Psychological and educational perspectives. New York: Macmillan.

Terman, L. (1925). Mental and physical traits of a thousand gifted children (Vol. I). Palo Alto, CA: Stanford University Press.

Tonemah, S.A. (1991). Philosophical perspectives of gifted and talented American Indian education. Journal of American Indian Education, 31(1), 3-9.

Torrance, E. P. (1978). Ways of discovering gifted Black children. In A.Y. Baldwin, G.H. Gear, & L.J. Lucito (Eds.), Educational planning for the gifted. Reston, VA: The Council for Exceptional Children.

Treffinger, D.J., & Renzulli, J.S. (1986). Giftedness as potential for creative productivity: Transcending IQ scores. Roeper Review, 8(3), 150-154.

Tyler-Wood, T., & Carri, L. (1993). Verbal measures of cognitive ability: The gifted low ses student's albatross. Roeper Review, 16(2), 102-105.

Udall, A., Reid, C., & Romanoff, B. (1995). Program for the gifted second grade assessment. Charlotte, NC: Program for the Gifted, Charlotte-Mecklenburg Schools.

U.S. Congress (1994). Improving America's Schools Act. Title X, Part B. Conference Report to Accompany H.R. 6. Report 103-761. September 28, 1994.

U.S. Department of Education (1993). National excellence: A case for developing America's talent. Washington, DC: U.S. Department of Education.

U.S. Department of Education (1994). Javits gifted and talented students education program: Grants projects abstracts, 1992-1993.

Van Tassel-Baska, J., Patton, J., & Prillaman, J. (1991). Gifted youth at risk: A report of a national study. Reston, VA: The Council for Exceptional Children.

Vernon, P.E., Adamson, G., & Vernon, D.F. (1977). The psychology and education of gifted children. Boulder, CO: Westview Press.

Walters, J. & Gardner, H. (1986). Crystallizing experiences: Discovering an intellectual gift. In R.J. Sternberg & J. Davidson (Eds.), Conceptions of giftedness. Cambridge: Cambridge University Press.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70(9), 703-713.

Wiggins, G. (1993a). Assessment: Authenticity, context, and validity. Phi Delta Kappan, 75(3), 200-214.

Wiggins, G. (1993b). Assessing student performance. San Francisco: Jossey-Bass.

Winner, E. (1996). Gifted children: Myths and realities. New York: BasicBooks.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), Review of Research in Education, 17, 31-74. Washington, DC: American Educational Research Association.

Wolf, D., LeMahieu, P.G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. Educational Leadership, 49(8), 8-13.

Worthen, B.R. (1993). Critical issues that will determine the future of alternative assessment. Phi Delta Kappan, 74(6), 444-48, 450-54.

PERSONAL COMMUNICATIONS

Ms. Liz Barnes, Team Leader, Javits Grant Projects, Office of Educational Research and Improvement, U.S. Department of Education. February and March 1997 phone calls regarding Javits grants to Charlotte-Mecklenburg, DISCOVER, and Montgomery County.

Dr. Howard Gardner, Professor, Harvard University. April 1996 remarks during the oral examination on the proposal to this dissertation; April 1997, memorandum on the draft of this dissertation.

Ms. Rebecca Garland, AG Consultant, Exceptional Children Division, North Carolina Department of Public Instruction. March 1997 phone conversation regarding North Carolina's regulation and financing of programs for the gifted.

Dr. C. June Maker, Professor, University of Arizona at Tucson. February 1997 phone conversation regarding DISCOVER III.

Dr. Aleene Nielson, Assistant Adjunct Professor, University of Arizona at Tucson. February 18, 1997 letter regarding my dissertation draft chapter on DISCOVER III.

Dr. Gary Orfield, Professor, Harvard University. April 1997 conversation about his research in the Montgomery County Public Schools.

Ms. Jean Paynter, Facilitator for Gifted and Talented Education, Maryland State Department of Education. February and March 1997 phone conversations regarding Maryland's regulation and financing of programs for the gifted.

Dr. Judith Rogers. Coordinator, DISCOVER III, University of Arizona at Tucson. March 1995 phone call regarding sites using the theory of multiple intelligences to identify students for gifted and talented programs.

Dr. Patricia O'Connell Ross, Executive Director of the Javits Program, Office of Educational Research and Improvement, U.S. Department of Education. April, 1995 phone call regarding Javits-funded sites using the theory of multiple intelligences.

Mr. Peter Rosenstein, Executive Director, National Association of Gifted Children, March 1997, regarding benefits accruing to youngsters who participate in programs for the gifted.

Dr. Jean Stahl, Gifted Education Specialist, Arizona Department of Education. February 1997 phone call regarding Arizona's regulation and financing of programs for the gifted.

Mr. Jack Stewart, Charlotte-Mecklenburg Chamber of Commerce. December 22, 1996 phone conversation about Charlotte-Mecklenburg's economic and business environment.

APPENDIX A
INTERVIEWEES

DISCOVER III

Name	Designer	Assessor	Principal/GT Teacher
Susan Bartley		x	x
Lorraine Etcitty			x
C. June Maker	x	x	
Carol Martin			x
Claudia MacArthur		x	
Aleene Nielson	x	x	
Lee Nelson		x	
Judith Rogers	x	x	

Additional transcribed data came from a meeting with Rogers.

CHARLOTTE-MECKLENBURG SCHOOLS' PSA

Name	Designer	Assessor	Principal/GT Teacher
Ty Fox		x	
Steve Houser		x	x
Mindy Passe	x		
Carol Reid	x		
Brenda Romanoff	x	x	
Anne Udall	x		
Becky Workman		x	x

Additional transcribed information came from a meeting with Romanoff; a meeting with Reid and Romanoff, and a meeting with Reid, Romanoff, Udall, Bob Algozzine, and several teachers of the gifted.

MONTGOMERY COUNTY'S MODEL PROGRAM

Name	Designer	Assessor	Principal/Teacher
Brian Bartels		x	
Karen Bulman		x	x
Carol Hylton		x	x
Debbie Leibowitz	x	x	
Pam Sobel			x
Waveline Starnes	x		
Barbara Williams		x	x

Additional transcribed information came from a meeting with Hylton, Leibowitz, and Donnelly Gregory; a meeting with Jean Barton, Bartels, and Starnes, and a meeting with Bartels, Hylton, Prue, Leibowitz, Starnes, and Ginny Tucker, acting director of Enriched and Innovative Services.

APPENDIX B
Questionnaire for Javits Program Assessment Designers*

INTRODUCTION:

Thank you for agreeing to speak with me today about the assessments you have developed to identify youngsters for gifted education. As you may know, I'm conducting this interview with you and others to collect data for my doctoral dissertation at the Harvard Graduate School of Education. I'm also a research coordinator at Harvard Project Zero, where I work with Professor Howard Gardner. For my doctoral research, I am trying to understand how Gardner's theory of multiple intelligences is being used to identify elementary students for gifted and talented education -- especially students from groups that are usually underrepresented in G&T programs. I'm also curious to know about any outcomes to date associated with using these identification procedures. To help me understand these issues I'm focusing on the Javits-funded programs that have used MI.

To learn more about these issues, I've developed an interview guide. It has four sections. The first deals with background about you, the Javits program, the district or school you are working in, and the change over to the assessments you are now using. The second asks about the assessment activities and procedures that you now use for identifying students for G&T. The third asks about how people interpret the data obtained from these instruments and procedures. The last section asks about outcomes or changes mostly in the number and kinds of students selected for gifted education since the Javits-funded effort was begun.

Do you have any questions?

Re confidentiality: It's important to mention that I don't plan to hide the identification of the Javits program, since other published materials name the program, describes its identification approach, and the populations with which it works. However, if you would like, I do not need to identify you. If you as an individual want to remain unidentified, I would not use your name and I would mask your comments so that they would be hard to attribute to you. If you would like to remain unidentified, please let me know. [PAUSE] If you decide it is alright to identify you, but in the course of interviewing feel that a particular statement should not be attributed to you, please let me know. In that case, if that statement or idea were used I would present it in a way that would make it hard for anyone to attribute it to you.

Would it be alright to tape record this interview?

(*This is one of three interview guides; the others, for assessment administrators/ observers and for teachers of the gifted and talented, closely parallel this one.)

TRANSITION: OK, let's start with background

Background

Could you tell me about the district/schools you are working in?
- student demographics
- community demographics
- special programs
- guiding philosophies

Could you tell me a bit about your background?
How did you come to be associated with the Javits program?

How long have you been associated with the school/district?
What other roles have you had in the school/district?

Were you working in the school/district when the decision to change id procedures for G&T was made?

Could you describe what the process for identifying students for G&T was in the past for the district[s] you are working in?

What children were assessed under the old process?
Why?
Who decided whether a child was assessed for G&T?

What did the G&T program consist of during the time the earlier assessment process was in place?
Enrichment?
Separate schools/curriculum?

Are there any descriptions of the previous identification process, instruments, and the gifted and talented program that I might get copies of?

Could you describe what prompted the change in the previous process used to identify students for G&T?
What was the goal in changing the assessment procedure?
Where did these goals come from?
Who supported the change?
What, if any, resistance to the change in assessment was there?

How did you come to draw on MI as a vehicle for selecting students for G&T?
- what considerations influenced your adoption of new assessment strategy using MI [cost, philosophy, etc.]

To what extent have you drawn on Gardner's writings on assessment in adapting MI for the identification process?

Have other writings or research on assessment practices shaped your identification process?

TRANSITION: That's about all the questions I have about background. Do you think there is something I left out or should have asked about your background, the school/district you are working in, or the change to the G&T assessment you are now using?

II. This next set of questions is aimed at getting a sense of what the assessment tasks and procedures entail: the materials used by students and assessors, the physical space in which the assessment occurs, the kinds of interactions that occur among the people in the room, the number of children and adults present; the time frame over which it occurs.

Who among the children now participates in this identification process?
- on what basis?

What would a youngster participating in this identification process see and do?
-What materials is the youngster using?
-What is the youngster's physical surrounding while this is going on?
- who, if anyone else, is present in the room besides youngster and assessor?

How long a period of time would the child spend in the assessment process?

Is the assessment repeated?

Is the assessment videotaped or recorded in some fashion?

How are assessors trained to carry out assessments with students using these tasks?

What would the administrator see and do during the identification process?
- What exactly are the instruments that the administrator/ assessor uses to gather information about individual students?
- What does the administrator do when a youngster is having difficulty with a task or question?

What challenges does the administrator or assessor face in carrying out the identification process?

Are there copies of instruments, procedure guidelines or other literature related to the current identification process that I could get copies of?

TRANSITION. Those are the questions that I have about the tasks and procedures

used in the identification process. Are there things you think I should know or questions that I should have asked about the tasks and procedures?

THE THIRD SET OF QUESTIONS IS AIMED AT UNDERSTANDING HOW THE INFORMATION YOU COLLECT IN THE ASSESSMENT PROCEDURES GETS INTERPRETED OR EVALUATED.

Who evaluates the information obtained in the assessment procedure?
Why these people?
How are they trained?

Are the people who interpret the information collected during the identification process the same people who make the decisions about placement into G&T?
Who else participates in the decision-making process for placing children into G&T

Could you walk me through the process by which the information collected on each child gets evaluated? [In other words, how do evaluators go from information collected during the assessment process to determinations about whether a child should or shouldn't be placed in G&T?]

What gives you confidence that this process identifies children accurately?
What procedures does your program use to achieve reliability among the assessors or administrators?
- inter-rater reliability
- intra-rater reliability
- student performance over time/test-retest reliability

What if any procedures are used to help make sure that youngsters who are gifted are detected? [false negatives]

Is there anything else about the interpretation of assessment information that you think I ought to know or should have asked about?

TRANSITION
Now, I'd like to move onto outcomes, to get a sense of what changes have occurred with the new identification process.

What percent of children who are assessed under these methods are now identified as gifted?

How does this number compare with the percent of children identified as gifted under the previous system?

In what if any way has there been a change in the proportion of poor and minority students identified as gifted?

Do you have information about how the children identified under the new identification process are performing?

Information from teachers?
Information from standardized tests?
Other measures?

Is there data that compares these youngsters' performance to children selected for G&T by other procedures (e.g., standardized tests)?

What indicators do you have that youngsters identified under the new system are "gifted."

- concurrence by teachers?
- outside awards in areas identified as gifted (e.g., dance, music, drawing?

Are there other things with regard to outcomes from the identification process that you think I should have asked or should know?

I know this has been a very long interview, but if I have forgotten to ask you something, may I contact you again?

THANK YOU

APPENDIX C
CODING SCHEME

CONDITIONS
best work+
best work
best work-
children understand tasks+
children understand tasks
children understand tasks-
clear scoring procedures+
clear scoring procedures
clear scoring procedures-
observer reliability+
observer reliability
observer reliability-
observer training+
observer training
observer training-
domain-based+
domain-based
domain-based-
intelligence-fair+
intelligence-fair
intelligence-fair-
traditional 3 abilities+
traditional 3 abilities
traditional 3 abilities-

CONTEXT
curriculum-assessment link
classroom
culture
leadership
research effort
school
school district
state policy

EVALUATION CRITERIA
benefit of the doubt
classroom standard
consensus
future curriculum
gut reaction
openended
reality check
teacher reference
three-dimensional
unique

OUTCOMES
high stakes
low stakes
slow change
teacher change
who now assessed
who now identified
who now served

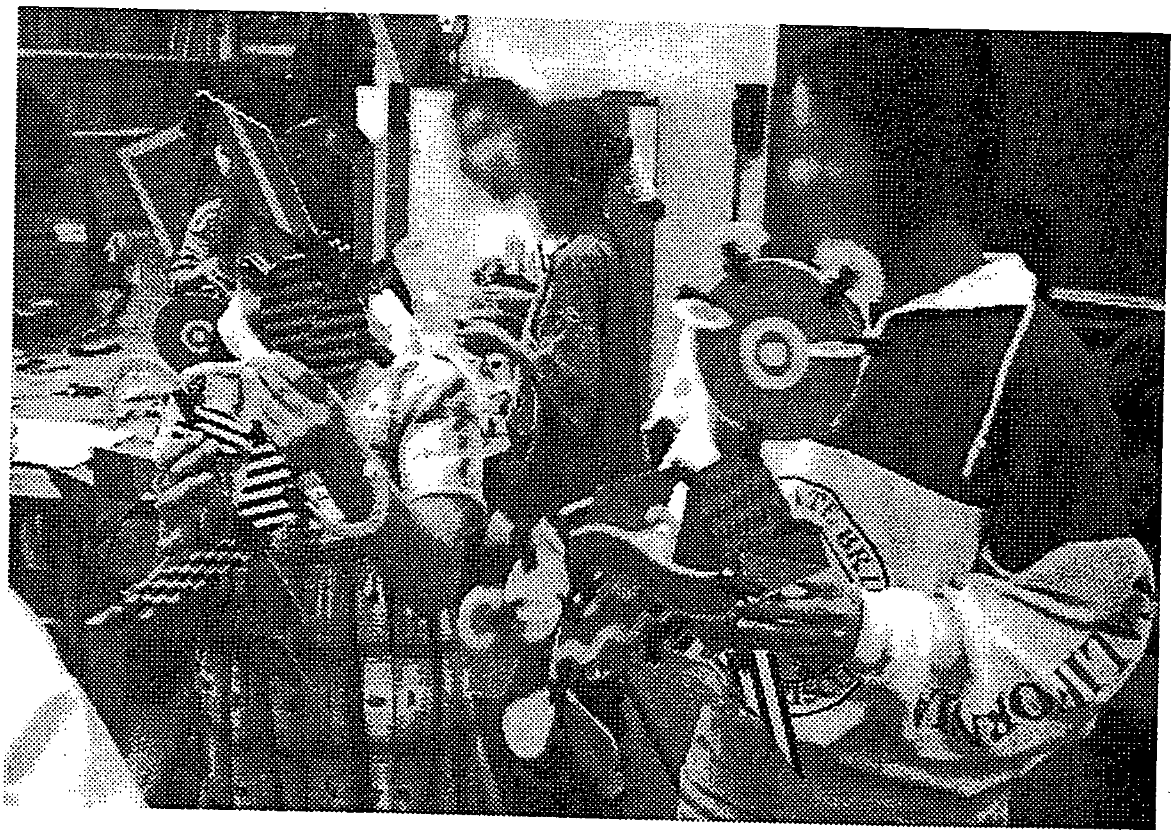
PROCEDURES
Instructions to students
INSTRUMENTS:
CMS observer checklist
CMS teacher checklist
DISCOVER checklist
DISCOVER obs. notes
DISCOVER personal interaction sheets
MCPS checklist

TASKS
categories
context
cms-linguistic
cms-logicalmath
cms-spatial
map
math booklet
mathfluency
mathsequence
mathstory
Pablo®
storytelling
storywriting
tangram

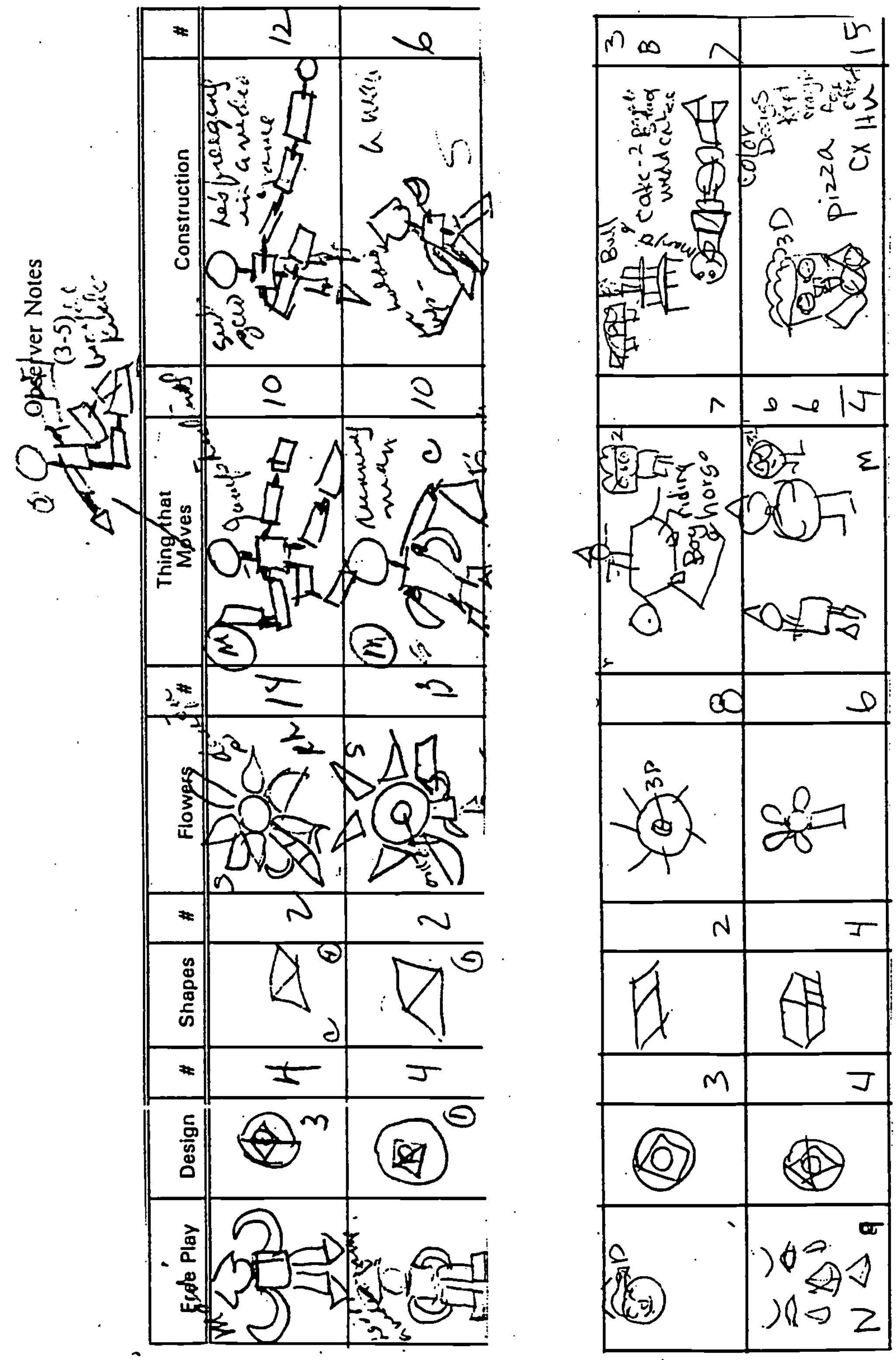
TESTS
CogAT
Iowa
MAT
Raven's
Standardized

OTHER
students' experiences
observers' role
equity
'g' versus MI
PADI
reliability of students' performance
validity

APPENDIX D
PABLO® MATERIALS



APPENDIX E
DISCOVER DOCUMENTATION OF PABLO®



BEST COPY AVAILABLE

APPENDIX F

CHARLOTTE'S SECOND GRADE CLASSROOM PORTFOLIO

Not Referred for Evaluation ☐

Charlotte-Mecklenburg Schools
Program For The Gifted
Second Grade Classroom Portfolio

Student Name ______________________ Classroom Teacher ______________________

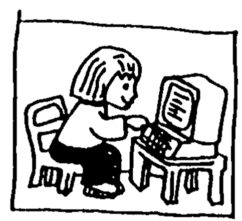
Spatial
This child loves to imagine real things, draw real representations, knows where everything is, and is fascinated with contraptions.

Likes...
drawing, building, designing, creating, visualizing, colors, videos, photos

Is Good At...
creating models, mapping, sculpting, imagining, puzzles

Learns By...
manipulating, imagining, creating art

Mathematical
A child with this strength loves to think conceptually, see patterns, figure out, question, wonder, and compute.

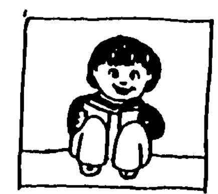


Likes...
numbers, patterns, relationships, formulas

Is Good At...
reasoning, logic, problem solving number games, computers

Learns By...
categorizing, collecting, working with patterns, brain-teasers

Linguistic
This child loves stories, jokes, riddles, tongue twisters, and trivia.



Likes...
words, reading, storytelling, writing, spelling

Is Good At...
writing & telling stories, poetry, vocabulary, memory

Learns By...
talking, listening, reading word games, acting

The following activities and books may be used with second graders to explore their intelligences.

Spatial Intelligence

These children think in images and pictures. They are often very aware of things in their environment. They like to draw, paint, make interesting designs, work with clay, colored construction paper, and fabric. They love jigsaw puzzles, reading maps, finding their way someplace new, and daydreaming. They have strong opinions about colors that go together, design, textures that are pleasing, and decorating. They are excellent at performing tasks that require "seeing with the mind's eye."

Dacta-Lego
Tessellation by Seymour
Visual Thinking by Seymour
Introduction to Tessellation by Seymour
Crazy Puzzles by Heye Concepts
Going Beyond Words by Mason
Perceptual Puzzle Blocks by Creative Publications
Building Thinking Skills (Midwest)
Pattern Games (checkers, chess, Rubik's Cube, tic-tac-toe)
Math Brainstorming (Good Apple)
Graphical Representation Games (Pictionary, connect the dots)
Imagining Games (Jigsaw puzzles, "what's wrong with the picture")
Scavenger Hunts
Map Reading

Linguistic Intelligence

These children have highly developed verbal skills and think by carrying on a conversation in their mind. They usually like reading, playing word games, making up poetry and stories, getting into involved discussions, debate, formal speaking, creative writing, and telling jokes. They tend to be precise in expressing themselves, love learning new words, do well in writing, and have a high reading comprehension.

Word Games
Fantasy Fairy Tales (Good Apple)
Scrabble
Learning Delights by Glasscock
Computer Software such as *Carmen Sandiago*
Mastering Reading Through Reasoning by Whimbex
Crossword Puzzles
The Flying Circus of Physics With Answers
Word Guessing Games such as hangman
Thinking Is The Key by Johnson
Impromptu Speaking Games
Brainstorming by Dickinson
Riddles, Puns, Jokes
Make Learning Fun (OM)

Logical/Mathematical

Children who show this preference think conceptually and abstractly, and are able to see patterns and relationships that others often miss. They like to experiment, solve puzzles and other problems, ask cosmic questions, and think. They generally enjoy working with numbers and math formulas/operations. They love the challenge of a complex or involved problem. They tend to be systematic and analytical, and they always have a logical rationale or argument for what they are doing or thinking.

Magic Squares
Math for Smarty Pants
Pattern Blocks
Math for Girls
Graphing lessons
Spaces by Dale Seymour
Probability
Logic Problems for Primary People by Seymour
Attribute Blocks
Collection of Math Lessons by Burns & Tank
Pre-Algebra Kit
Creative Problem Solving by Lenchner
Metric Kit
Mathematical Mystery Tour by Wahl
Hands-on Story Problems
Math for Math Lovers (Sunburst)
Toothpick Puzzles
Pentaminios
Tops Kit

Mathematics In Action math text, pages:
4, 8, 65, 78, 139, 175, 182, 199, 205, 258, 282-284, 311, 345, 311, 411

BEST COPY AVAILABLE

CMS Program for the Gifted
Referral Form: Multiple Intelligences

Student Name: ____________________ Referred by____________________ Relationship to Student________________

Directions: Please think about this student. Check the degree of behavior for each example that applies.

	Not Evident	Evident	Strongly Evident	Always Evident
Linguistic Intelligence				
• is an avid reader				
• enjoys telling detailed and expressive stories				
• enjoys writing and/or reading				
• persuasive and precise in expressing self				
• enjoys and can create such things as puns, riddles, metaphors, & analogies				
• uses words to create vivid images or emotions				
• relates experiences in vivid detail through speaking or writing				
• other evidence of linguistic intelligence:				
Logical - Mathematical Intelligence				
• excellent at finding and remembering patterns				
• can easily remember formulas and strategies				
• highly observant				
• unusual skill in taking apart and reassembling things				
• loves to sort objects and ideas into categories				
• enjoys complex number problems and can solve them				
• sees many different and/or unusual ways to solve problems				
• challenges other people's thinking processes and decisions				
• other evidence of logical mathematical intelligence:				
Spatial Intelligence				
• likes to draw, doodle, copy, trace, and draw freehand; drawing reflects complexity				
• enjoys drawing, painting, working with clay, constructing models, designs are complex				
• easily designs, assembles, constructs, and/or manipulates forms and shapes				
• loves puzzles, legos, pictionary, chess and shows exceptional talent in those areas				
• can easily imagine how an object will appear from a different angle				
• solves problems efficiently by creating mental images				
• other evidence of spatial intelligence:				
Multiple Intelligence/Problem-Solving				
• is persistent in behavior such as questioning or task commitment				
• invents new ways to solve problems				
• solves problems quickly				
• shows or verbalizes enjoyment of a challenging task				
• shows evidence of logical thought				
• completes tasks independently				
• other evidence of extraordinary intelligence:				

Please return to the school based AG teacher

APPENDIX G
CHARLOTTE'S "YELLOW CARD"

CMS PROGRAM FOR THE GIFTED
PROBLEM-SOLVING BEHAVIOR OBSERVATION CARD

Decision Regarding Placement of ____________________

Date: _____/_____/_____

We, the Observation Team, have reviewed all the documentation and

☐ **Do recommend place in the Encounter Resource Program**

☐ **Do not recommend place in the Encounter Resource Program**

FINAL SUMMARY OF OBSERVED BEHAVIORS

Area of Intelligence	Not Evident	Evident	Strongly Evident	Always Evident
Spatial				
Lingustic				
Logical Mathematical				

Observation Team

Comments: ____________________

PG Teachers: ____________________

Team Members: ____________________

Principal: ____________________

Matrix Analogies Test (Short Form)

	YEAR	MONTH	DAY
Date Tested			
Date of Birth			
Chronological Age			

Total of Correct Answers: __________

Stanine Score: __________

Age = ____years, ______months

Observed Behaviors: Linguistic Problem-Solving
Storytelling and Storywriting

2nd Grade

Student________________ Observer________________ Date______

Check the Points That Apply:	Storytelling	Storywriting
Story has a clear introduction and conclusion.		
Story has a clear sense of place.		
Plot includes action.		
Plot has a logical sequence and complete thoughts.		
Story includes comparisons and similarities or differences.		
Child uses complete sentences.		
Child uses dialogue or conversation.		
Descriptions are detailed, creating a visual image.		
Vocabulary includes complex or sophisticated words.		
Vocabulary includes adjectives or adverbs.		
Child needed no prompting, language flows easily and fluently.		
Child shows enjoyment of the task.		
Child shows involvement in task... is not easily distracted.		
Child's voice shows expression and feeling.		
Story includes emotion and feeling.		
Child uses humor.		
Child uses props effectively to dramatize the story.		
Other:		
Other:		

Comments:________________________________

2nd Grade

Observed Behaviors- Linguistic Problem Solving

Context and Categories

Student____________________ Observer________________________ Date__________

Context	Categories	
☐	☐	Needed clarification of directions
☐	☐	Needs assistance with reading of items
☐	☐	Needs assistance with understanding

Context

___/5__ Other____________________

Categories

Titles Items

_/6_ _/6_ Creative________ Open-Ended ______

Comments____

Observed Behaviors: Spatial Problem Solving

	MAP	NEEDS PROMPTING
____	Use of road names ____________________	____
____	Use of left/ right directions ____________________	____
____	Use of place names ____________________	____
____	Use of cardinal directions ____________________	____
____	Completes all errands ____________________	____
____	Avoids disasters ____________________	____
____	Plans route before tracing ____________________	____
____	Reaches final destination (Home) ____________________	____
____	Gives divergent responses ____________________	____
____	Takes logical route ____________________	____
____	Expeditious use of time ____________________	____

OBSERVED BEHAVIORS: MATHEMATICAL PROBLEM SOLVING
MATH BOOKLET AND STORY MATH

2nd Grade

Student ______________________________

Observer ______________________________ Date ______________

☐ Uses arithmetic correctly most of the time	☐ Appears to utilize math strategies rather than guessing the correct answer	☐ Works easily and quickly throughout
☐ Answers to open-ended math questions show complexity	☐ Is alert and attentive to story math clues	☐ Solves complex problems quickly
☐ Develops an unusual written or oral problem-solving strategy	☐ Easily remembers patterns to solve story math problems	☐ Persists on tasks that are difficult
☐ Asked a question or made a comment that shows high-level thinking	☐ Enjoys complex, challenging problems	☐ Seems excited, absorbed in task
	☐ Solves all problems without frustration	☐ Increases in motivation or enjoyment as problems increase in open-endedness
		☐ Doesn't want to stop

_____ RECEIVED HELP; IF SO, COMMENT BELOW *STRATEGIES/NOTES*

_/4_O ANSWERS TO PART I. SEQUENCES ______________________________

_/4_O ANSWERS TO PART II. FUNCTIONS ______________________________

_/4_O ANSWERS TO PART III. NUMBER LOGIC ______________________________

_____ ANSWERS TO PART IV. FLUENCY SHOW USE OF A STRATEGY (11-1=10, 12-2=10 ETC.) ______________

_____ ANSWERS TO PART IV. FLUENCY USE MORE THAN ONE FORMULA (X, /, -, +) ______________

_____ ANSWERS TO PART IV. FLUENCY ______________________________

_/3_ ANSWERS TO STORY MATH ______________________________

Comments: ______________________________

____

OBSERVED BEHAVIORS: SPATIAL/LOGICAL PROBLEM SOLVING
TANGRAMS

Student ______________________

Observer ______________________ Date ______________

☐ Completed Square   Number of Pieces in Square ______   ☐ Helped Others

☐ Received Help fonn Observer or Classmate - Note which clues were offered:

☐ You have enough pieces to do this page
☐ You may have to use more than one piece to make some of the shapes
☐ You may need to take the pieces off and start over again
☐ Use these pieces here
☐ Find a place to use this piece /these pieces
☐ Make this puzzle a different way

Time Started __________   Time Stopped __________

☐ Makes unusual square
☐ Makes square in a logical way
☐ Reworks pieces without prompting to complete page
☐ Manipulates pieces to check sizes
☐ Selects correct pieces instinctively without manipulation
☐ Combines shapes, experiments
☐ Makes puzzles in more than one way
☐ Takes pieces off
☐ Substitutes pieces without clues
☐ Uses logical strategy for adding or substituting pieces
☐ Incorporates clues and new information

☐ Works on several constructions at one time
☐ Check sizes of pieces by putting one on top of another
☐ Shows humor in interactions with others
☐ Works independently
☐ Works continuously
☐ Works easily and quickly throughout
☐ Competes with others
☐ Encourages others
☐ Works complex problems quickly
☐ Persists on tasks that are difficult
☐ Seems excited and absorbed in task
☐ Doesn't want to stop

PAGES	ORDER	TIME
1.	______	______
2.	______	______
3.	______	______
4.	______	______
5.	______	______
6.	______	______

☐ Completed all pages correctly without specific clues   Completed only through page ______

Comments: ______________________

OBSERVED BEHAVIORS: SPATIAL PROBLEM SOLVING
PABLO

Student ____________________

Observer ____________________ Date ____________

☐ Received help ☐ Copied idea of another student

☐ Clearly resemble what child intends to make
☐ Realistic object (objects, animals, environment)
☐ Product reflects an environment removed form child's daily live
☐ 3-Dimensional
☐ Variety of constructions
☐ Complex (# of pieces, # of connectors, interesting use of connectors)
☐ Other ____________
☐ Other ____________

☐ Original or unusual solution
☐ Moveable parts
☐ Symmetrical
☐ Asymmetrical with attention to design/detail
☐ Makes many creations
☐ Attends to Design on pieces
☐ Attends to Detail
☐ Substitutes pieces for a different piece
☐ Stretches boundaries of task

☐ Directs the spatial component of the group
☐ Hitchhikes on ideas of others without duplicating
☐ Works easily and quickly throughout
☐ Persists on tasks that are difficult
☐ Seems excited, absorbed in task
☐ Increases in motivation or enjoyment as problems increase in open-endedness
☐ Doesn't want to stop

Free Play	Shapes	Animal	Buildings	Something Moveable	PCS	Free Construction	PCS
					CNT		CNT

PCS = pieces CNT = connectors

Comments: ____________________

BEST COPY AVAILABLE

APPENDIX H
THE MONTGOMERY COUNTY GRID
GIFTED AND TALENTED DATA RECORDING FORM
Grade 2

School ________ Teacher ________ Date ________

Race/
Ethnic Code:
1-No. Amer. Ind.
2-Asian
3-Black
4-White
5-Hispanic

Student	Race (See chart)	Sex (M or F)	Parent Nomination/Permission (Y or N)	Kough/DeHaan Checklist # of characteristics	Others School/Community	Raven Matrices Standard Raw Score	Renzulli-Hartman (Points)			Test of Cognitive Skills				Other Assessment Data	Selection Decision (Y or N)	Comments
							Learning	Motivation	Creativity	Sequences	Memory	Analogies	Verbal Reasoning			

BEST COPY AVAILABLE

APPENDIX I

THE MONTGOMERY COUNTY READY REFERENCE

GRADE 2 SCREENING INSTRUMENTS
GLOBAL AND SPECIFIC COMBINED

SOURCE OF INFORMATION	RECOMMENDED INSTRUMENT	USE	CRITERION
1. Parent	MCPS Parent Nomination Form	Solicit from all parents	Return of form to school is one indicator.
2. Other Sources			
a. In-School Staff	MCPS Staff/Community Nomination Form	Circulate among school staff	Name listed on form is one indicator.
b. Community	MCPS Staff/Community Nomination Form	Community representatives determined by principal and school committee	Name listed on form is one indicator.
3. Classroom Teacher	a. Gough/DeBoen Checklist	Complete for current class	3 or more behavioral characteristics is one indicator.
	*b. Renzulli-Hartman Behavioral Rating Scales	Complete for any student who meets one or more criterid	Learning-24 or more Motivation-24 or more Creativity-25 or more Any 2 of these scores is one indicator.
*4. Standard Progressive Matrices (Raven)	Standard Raven	Administer to all students	Achieving the raw score as listed in the chart below is one indicator. These raw scores represent the 75%ile on MCPS norms and above the 90%ile on national norms. (1984)

Standard Raven

Age	Raw Score	*
7	33	(31)
7½	35	(24)
8	37	(30)
8½	39	(33)
9	41	(33)

*NOTE: Scores in parentheses represent performance at the 90%ile on the old Irish norms. Carefully consider a score between the old and new criterion. This is especially important for black and Hispanic students whose scores may be an underestimation of their true ability.

SOURCE OF INFORMATION	RECOMMENDED INSTRUMENT	USE	CRITERION
*5. Test of Cognitive Skills (TCS) Level 1	Subtests: Sequences Memory Analogies Verbal Reasoning	Administer to all students	Meeting the criterion on two of the four subtests is one indicator. See chart below for specific criteria. These raw scores represent 8th and 9th stanines.

TCS Level 1

	Raw Score
Sequences	19
Memory	15
Analogies	19
Verbal Rsng.	18

NOTE: Scores within 2-3 points of the criterion should be carefully considered. This test has a low ceiling for bright students.

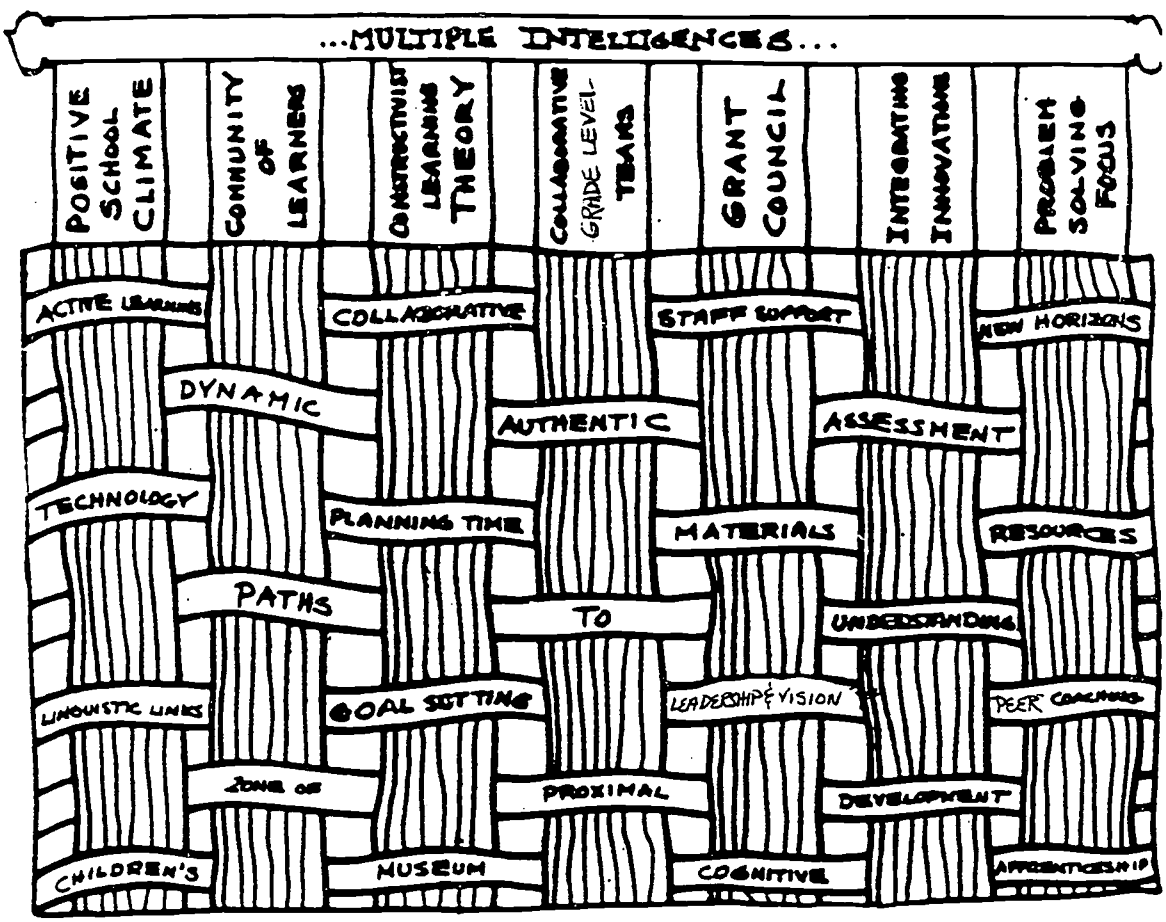
SOURCE OF INFORMATION	RECOMMENDED INSTRUMENT	USE	CRITERION
*6. Other Performance Data	a. Rdg/Math Levels Two or more yrs. above grade level	Review folders	One indicator
	b. Acceleration to a higher grade		One indicator
	c. Other evidence of outstanding performance		One indicator

*See page 13 for implementing professional decision making and committee review of this screening data. At least one of the three indicators must be a starred item.

BEST COPY AVAILABLE

JACOB JAVITS GRANT
Division of Enriched & Innovative Instruction

CLASSROOM MULTIPLE INTELLIGENCES TAPESTRY
Montgomery Knolls Elementary School



BEST COPY AVAILABLE

APPENDIX K
THE MONTGOMERY COUNTY MI CHECKLIST

CHECKLIST FOR IDENTIFYING LEARNING STRENGTHS 1995/96

Child's name________________ Teacher__________
Grade______

Write the number (1 - 5) that most closely represents your overall observations of this child in each intelligence. You may check any behavior(s) you feel is (are) particularly strong for that child. Please add any comments you believe will help another teacher plan for this child.

1 = You have not observed these behaviors.
2 = You have occasionally observed them.
3 = You have usually observed them.
4 = You almost always or always observed them.
5 = No opportunity to observe these behaviors.

F = Fall, S = Spring

F S *LINGUISTIC* ______ (overall rating) F ___ S ___

___ ___ 1. Enjoys word play; chooses to memorize and recite poems, tongue twisters, puns, riddles, etc.
___ ___ 2 Starts conversations or discussions on his/her own.
___ ___ 3. Expresses ideas easily either orally or in writing. Is a good storyteller or writer.
___ ___ 4 Can give several meanings when describing an object or idea (e.g. how an object looks; how it's used).
___ ___ 5. Remembers and describes new ideas.
___ ___ 6. Readily verbalizes background knowledge and factual information.
___ ___ 7. Asks many questions.
___ ___ 8. Talks through problems; explains solutions.
___ ___ 9. Shows verbal ability in English, considering another language is used in the home.
___ ___ 10. Uses advanced vocabulary for age.

LOGICAL-MATHEMATICAL ______ (overall rating) ___ ___

___ ___ 1. Chooses to play or work with number activities.
___ ___ 2. Finds number patterns or geometric patterns in the environment (e.g., tiles, flowers, leaves).
___ ___ 3. Joins smaller ideas into larger ones.
___ ___ 4. Can provide specific examples to support a generalization.
___ ___ 5. Finds ways to work through an unfamiliar number problem using own plan or strategy.
___ ___ 6. Is able to plan or describe steps or events in order.
___ ___ 7. Groups objects and ideas in a variety of ways; finds similarities and differences.
___ ___ 8. Uses a systematic approach to problem solving.
___ ___ 9. Assembles puzzles with skill and enjoyment.

BODILY-KINESTHETIC ______ (overall rating) ___ ___

___ ___ 1. Chooses motor skills (e.g. skipping, balancing, jumping).
___ ___ 2. Mirrors or repeats movements easily.
___ ___ 3. Readily masters hand (clap) patterns or steps.
___ ___ 4. Develops large muscle (gross motor) skills easily (e.g. roller skating, jumping rope).
___ ___ 5. Develops small muscle (fine motor) skills easily (e.g. tying shoes before kindergarten, draws unusually well for age).
___ ___ 6. Tries to master a new physical skill independently.
___ ___ 7. Prefers to touch and explore the shape of objects in order to learn about them.

Montgomery County Public Schools
Gifted Unit
850 Hungerford Drive
Rockville, Maryland 20850

GMP - Pine Crest E.S.
201 Woodmoor Drive
Silver Spring, MD 20901

5

SPATIAL _______ (overall rating)

1. Chooses to express ideas through visual media or through interactions with objects in the environment.
2. Constructs and designs visual patterns
3. Puts things together imaginatively to form a construction (e.g. collage, sculpture)
4. Shows an understanding of physical perspective.
5. Takes things apart and can put them back together (e.g. puzzle or mechanical objects).
6. Can organize and group objects.
7. Shows artistic appreciation, responds to color, line, texture.
8. Carefully plans use of space on paper.
9. Puts relevant detail in drawings.

INTERPERSONAL _______ (overall rating)

1. Eager participant in group activities.
2. Initiates or makes offers of peer tutoring.
3. Meets own needs through adults and other people.
4. Expresses feelings to others.
5. Shows leadership; organizes activities including other children.
6. Chosen by others to help or join a group.
7. Easily builds relationships with others.
8. Shows strong sense of fairness in the interest of the group.

INTRAPERSONAL _______ (overall rating)

1. Self motivated, independent and resourceful.
2. Accepts ownership for one's behavior.
3. Self confident.
4. Empathizes with other children.
5. Has sense of humor.
6. Can laugh at oneself.
7. Sticks to one's beliefs.
8. Takes risks.
9. Concentrates on topics or tasks.
10. Plays creatively.
11. Persistent in self selected activity.

MUSICAL _______ (overall rating)

1. Chooses musical activities.
2. Reproduces a newly heard melody or rhythm.
3. Composes rhythms, patterns, melodies.
4. Sings on key.
5. Identifies musical instruments heard in a musical composition.
6. Plays musical selections by ear.
7. Sings or hums melodically during independent activities.
8. Experiments with objects to create different sounds.

Comments:

VITA

Mindy Laura Kornhaber

1975-1978	Boston University Boston, Massachusetts	B. Mus. August 1978
1979-1982	Writer New York, New York	
1983-1986	Administrator Columbia University New York, New York	
1987-1988	Graduate School of Education Harvard University	Ed.M. June 1988
1988-1989	Research Assistant Project Zero Graduate School of Education Harvard University	
1989-1997	Doctoral Candidate Graduate School of Education Harvard University	
1990-	Researcher Project Zero Graduate School of Education Harvard University	